

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2002, 17:39:12 ; Search time 218.16 Seconds  
(without alignments)  
789.279 Million cell updates/sec

Title: US-09-740-211-13\_COPY\_1000\_1700

Perfect score: 701  
Sequence: 1 atggagaccctactagtatg.....ctgcggcccgatgacagaag 701

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/6C.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	701	100.0	3 US-08-717-294-41	Sequence 41, Appl
2	701	100.0	4 US-09-470-618-14	Sequence 14, Appl
3	701	100.0	4 US-09-364-862-14	Sequence 14, Appl
4	701	100.0	5035 2 US-08-882-083-1	Sequence 1, Appl
5	701	100.0	5035 2 US-08-558-107-1	Sequence 1, Appl
6	701	100.0	5035 3 US-09-243-539-1	Sequence 1, Appl
7	701	100.0	6939 1 US-08-276-594A-1	Sequence 1, Appl
8	701	100.0	7056 1 US-08-121-202-1	Sequence 1, Appl
9	701	100.0	8241 6 5171844-1	Patent No. 5171844
10	701	100.0	9567 1 US-08-366-851A-1	Sequence 1, Appl
11	701	100.0	9009 1 US-07-864-004B-3	Sequence 3, Appl
12	701	100.0	9009 1 US-08-251-937A-3	Sequence 3, Appl
13	701	100.0	9009 1 US-08-212-133A-1	Sequence 1, Appl
14	701	100.0	9009 1 US-08-474-503-1	Sequence 1, Appl
15	701	100.0	9009 2 US-08-670-707A-1	Sequence 1, Appl
16	701	100.0	9009 5 US-09-037-601-1	Sequence 1, Appl
17	701	100.0	9009 5 PCT-US93-03275-3	Sequence 3, Appl
18	701	100.0	9009 5 PCT-US94-13200-1	Sequence 1, Appl
19	701	100.0	9354 1 US-08-683-839B-2	Sequence 2, Appl
20	701	100.0	11933 4 US-09-470-618-13	Sequence 13, Appl
21	701	100.0	11933 4 US-09-364-862-13	Sequence 13, Appl
22	699.4	99.8	4629 2 US-08-484-891-7	Sequence 7, Appl
23	510.6	72.8	7493 1 US-08-212-133A-7	Sequence 7, Appl
24	510.6	72.8	7493 1 US-08-474-503-5	Sequence 5, Appl
25	510.6	72.8	7493 1 US-08-670-707A-5	Sequence 5, Appl
26	510.6	72.8	7493 2 US-09-037-601-5	Sequence 5, Appl
27	510.6	72.8	7493 5 PCT-US94-13200-5	Sequence 5, Appl

28	506.2	72.2	7032 4 US-09-324-867-1	Sequence 1, Appl
29	491.4	70.1	4334 2 US-08-670-707A-38	Sequence 38, Appl
30	491.4	70.1	4334 4 US-09-037-601-38	Sequence 38, Appl
31	491.4	70.1	6402 2 US-08-670-707A-36	Sequence 36, Appl
32	491.4	70.1	6402 4 US-09-037-601-36	Sequence 36, Appl
33	431	61.5	1623 1 US-08-121-202-3	Sequence 3, Appl
34	398.6	56.9	4451 3 US-08-717-294-42	Sequence 42, Appl
35	62.4	8.9	6909 2 US-08-804-196-1	Sequence 1, Appl
36	62.4	8.9	6909 2 US-08-658-340-1	Sequence 1, Appl
37	62.4	8.9	6909 2 US-08-746-111-26	Sequence 26, Appl
38	61	8.7	6585 3 US-08-746-111-4	Sequence 4, Appl
39	58.6	8.4	1130 1 US-07-864-004B-1	Sequence 1, Appl
40	58.6	8.4	1130 1 US-08-251-937A-1	Sequence 1, Appl
41	58.6	8.4	1130 1 US-08-212-133A-5	Sequence 5, Appl
42	58.6	8.4	1130 1 US-08-474-503-3	Sequence 3, Appl
43	58.6	8.4	1130 2 US-08-670-707A-3	Sequence 3, Appl
44	58.6	8.4	1130 4 US-09-037-601-3	Sequence 3, Appl
45	58.6	8.4	1130 5 PCT-US93-03275-1	Sequence 1, Appl

## ALIGNMENTS

```

RESULT 1
: US-08-717-294-41
: Sequence 41, Application US/08717294
: Patent No. 6114148
:
: GENERAL INFORMATION:
: APPLICANT: SEED, BRIAN
: APPLICANT: HAAS, JURGEN
: TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF
: TITLE OF INVENTION: PROTEINS
: NUMBER OF SEQUENCES: 110
: CORRESPONDENCE ADDRESS:
: ADDRESS: Clark & Elbing LLP
: STREET: 176 Federal Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08-717,294
: FILING DATE: 20-SEP-1996
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Elbing, Karen L
: REGISTRATION NUMBER: 35,238
: TELEPHONE/DOCKET NUMBER: 00786/345001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-428-0200
: TELEFAX: 617-428-7045
:
: TELEEX:
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4670 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
:
: US-08-717-294-41
:
: Query Match 100.0%, Score 701, DB 3, Length 4670,
: Best Local Similarity 100.0%, Pred. No. 4.5e-213,
: Matches 701, Conservative 0, Mismatches 0, Indels 0, Gaps 0:

```

```
QY 1 attgagccctactagatagagaaggagtcgccaaggaagaacacagacttg 60
DB 600 ATTGAGCCCTACTAGTATGTAAGGAAGGAGTCTGGCCCAAGAAAAGACAGACTTG 659
QY 61 cacaattatatactacttttctgtatttgaagaaggaaagtggcactcagaaca 120
DB 660 CACAATTATATACTACTTTTCTGTATTTGATGAAGGAAAAGTTGGCACTCAGAAAACA 719
QY 121 aagaactccttgatgagagataagatgctgcatctctcgccgctcgctaaatgac 180
DB 720 AAGAACTCCTTGATGAGAGATAGGATGCTGCTGCTCGGCTTGCCCTTAATATGACAC 779
QY 181 acagtaaatggttaataaacaagtcctcgcaagtcgattgagatgacacagaataca 240
DB 780 ACAGTAATGTTATGTAACAAGTCTCTGCCAGTCTGATTGATGATGCCACAGGAATACA 839
QY 241 gtctattgcatgttgatggaatggcaccacacctcggaagtgcactcaatatctcgaa 300
DB 840 GTCTATTGGCATGTGATGGAAATGGCCACACTCTGGAAGTGCATCAATATTCCTCGAA 899
QY 301 ggtcacacatttctgtgaggaaccatcgccagcgctccttggaatatctcgccaatact 360
DB 900 GGTCAACATTTCTGTGAGGAACCATGCGCAGGCGTCTTGGAATCTCGCCATTAAT 959
QY 361 ttccctacgtctcaaacactcttgatggaaccttgacaagtttctactgtttgtcatatc 420
DB 960 TTCTCTACTGCTCAACACTCTTGATGAGACCTTGACACTTTCATCTGTTTTGTCAATATC 1019
QY 421 tcttcccaacaacatgatgcatggaagcttattgcaaaagtgaagcgttcacaaagaa 480
DB 1020 TCTTCCCAACCATGATGATGAGAGCTTATGTCAAAGTGAAGCTGTCCAGAGGAA 1079
QY 481 ccccaactcgaatgaaaaataatgaagaagcgaagactatgatgatcttactgat 540
DB 1080 CCCCACCTACGATGATAAATAATGAAGACGGAAGACTATGATGATGATCTTACTGAT 1139
QY 541 tctgaatgagatgtgtaagtttatgatgcaactctccttcttataccaattcgc 600
DB 1140 TCTGAATGATGATGTTGCTGGTTTATGATGATGATCAACTCTCTTCTTATCCAAATTTGCG 1199
QY 601 tcaattgccaagaagcatcctaacttgggtacattacattgctgctggaagaagaagac 660
DB 1200 TCAGTTGCCAAGAAGCATCTTAAACTTGGGTACATTATTCATTTGCTGGAAGAGAGGAC 1259
QY 661 tgggaactatgctcccttagtctcgcccccgatgacagaag 701
DB 1260 TGGGACTATGCTCCTTAGTCTCTCGCCCCCGATGACAGAAG 1300

RESULT 2
US-09-470-618-14
; Sequence 14, Application US/09470618
; Patent No. 6200560
; GENERAL INFORMATION:
; APPLICANT: Couto, Linda B.
; APPLICANT: Colosi, Peter C.
; TITLE OF INVENTION: Adeno-Associated Vectors for Expression of Factor VIII
; FILE REFERENCE: AVigen-04082
; CURRENT FILING DATE: 1999-04-08
; EARLIER FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: 09/364,862
; EARLIER FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: 60/125,974
; EARLIER FILING DATE: 1999-03-24
; EARLIER APPLICATION NUMBER: 60/104,994
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 4999
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-470-618-14

Query Match      100.0%; Score 701; DB 4; Length 4999;
Best Local Similarity 100.0%; Pred. No. 4,7e-213;
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgagccctactagatagagaaggagtcgccaaggaagaacacagacttg 60
DB 989 attgagccctactagatagagaaggagtcgccaaggaagaacacagacttg 1048
QY 61 cacaattatatactacttttctgtatttgaagaaggaaagtggcactcagaaca 120
DB 1049 cacaattatatactacttttctgtatttgaagaaggaaagtggcactcagaaca 1108
QY 121 aagaactccttgatgagagataagatgctgcatctctcgccgctcgctaaatgac 180
DB 1109 aagaactccttgatgagagataagatgctgcatctctcgccgctcgctaaatgac 1168
QY 181 acagtaaatggttaataaacaagtcctcgcaagtcgattgagatgacacagaataca 240
DB 1169 acagtaaatggttaataaacaagtcctcgcaagtcgattgagatgacacagaataca 1228
QY 241 gtctattgcatgttgatggaatggcaccacacctcggaagtgcactcaatatctcgaa 300
DB 1229 gtctattgcatgttgatggaatggcaccacacctcggaagtgcactcaatatctcgaa 1288
QY 301 ggtcacacatttctgtgaggaaccatcgccagcgctccttggaatatctcgccaatact 360
DB 1289 ggtcacacatttctgtgaggaaccatcgccagcgctccttggaatatctcgccaatact 1348
QY 421 tcttcccaacaacatgatgcatggaagcttattgcaaaagtgaagcgttcacaaagaa 480
DB 1409 tcttcccaacaacatgatgcatggaagcttattgcaaaagtgaagcgttcacaaagaa 1468
QY 481 ccccaactcgaatgaaaaataatgaagaagcgaagactatgatgatcttactgat 540
DB 1469 ccccaactcgaatgaaaaataatgaagaagcgaagactatgatgatcttactgat 1528
QY 541 tctgaatgagatgtgtaagtttatgatgcaactctccttcttataccaattcgc 600
DB 1529 tctgaatgagatgtgtaagtttatgatgcaactctccttcttataccaattcgc 1588
QY 601 tcaattgccaagaagcatcctaacttgggtacattacattgctgctggaagaagaagac 660
DB 1589 tcaattgccaagaagcatcctaacttgggtacattacattgctgctggaagaagaagac 1648
QY 661 tgggaactatgctcccttagtctcgcccccgatgacagaag 701
DB 1649 tgggaactatgctcccttagtctcgcccccgatgacagaag 1689

RESULT 3
US-09-364-862-14
; Sequence 14, Application US/09364862
; Patent No. 6221349
; GENERAL INFORMATION:
; APPLICANT: Couto, Linda B.
; APPLICANT: Colosi, Peter C.
; TITLE OF INVENTION: ADENO-ASSOCIATED VECTORS FOR EXPRESSION OF FACTOR VIII
; FILE REFERENCE: BY TARGET
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: AVIGEN-03743
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: 60/125,974
```

```

1  EARLIER FILING DATE: 1999-03-24
2  EARLIER APPLICATION NUMBER: 60/104,994
3  EARLIER FILING DATE: 1998-10-20
4  NUMBER OF SEQ ID NOS: 14
5  SOFTWARE: PatentIn Ver. 2.0
6  SEQ ID NO 14
7  LENGTH: 4999
8  TYPE: DNA
9  ORGANISM: Artificial Sequence
10 FEATURE:
11 OTHER INFORMATION: Description of Artificial Sequence: Synthetic
12 US-09-364-862-14

```

Query Match	100.0%	Score 701	DB 4	length 4899
Best Local Similarity	100.0%	Pred. No.	4	7e-213
Matches 701	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

OY	1	attgagcccttactagatgtatgtaaagaaggaaatctgycgcaagaagaagaacacacattg	60
Db	909	atctgagcccttaccacaaatgaatgaagaagaaggagtctcgccaagaagaagaacacgacttg	1048
OY	61	cacaaattatactactcttttgcctgtattctgtatgtaaagggaaaagtctgcactcgaaca	120
Db	1049	cacaatttatactactcttttgcctgtattctgtatgtaaagggaaaagtctgcactcgaaca	1108
OY	121	aagaatcccttgatgcaggatgaaggatgtgatctgcgtccggcgcttgccctaaatgcac	180
Db	1109	aagaatcccttgatgcaggatgaaggatgtgatctgcgtccggcgcttgccctaaatgcac	1168
OY	181	acagtcacatggttatgtgaanaacaggtctctgcagcagctctgtatgtatgtgcacaggaatca	240
Db	1169	acagtcacatggttatgtgaanaacaggtctctgcagcagctctgtatgtatgtgcacaggaatca	1228
OY	241	gtctattggtcatgtgtatgtaaatggtggcacacacctccctgaagtctacatcaatttccctgaa	300
Db	1229	gtctattggtcatgtgtatgtaaatggtggcacacacctccctgaagtctacatcaatttccctgaa	1288
OY	301	gtctacacattctctgtgaaagaaacatcgcgcagcgtctctgtgaatatctgcgaataact	360
Db	1289	gtctacacattctctgtgaaagaaacatcgcgcagcgtctctgtgaatatctgcgaataact	1348
OY	361	tctccttactgtctaaacactctttgatgagacctgtgacagtttctactgttttgtatatac	420
Db	1349	tctccttactgtctaaacactctttgatgagacctgtgacagtttctactgttttgtatatac	1408
OY	421	tcttcccccaacaatgtgtgcacatgaaagctctatgtcaaaagtgaacagcgtctcagaagaa	480
Db	1409	tcttcccccaacaatgtgtgcacatgaaagctctatgtcaaaagtgaacagcgtctcagaagaa	1468
OY	481	cccccaactacgaatgaanaaataatgaagaagcgggaagactatgatgatattactgcat	540
Db	1469	cccccaactacgaatgaanaaataatgaagaagcgggaagactatgatgatattactgcat	1528
OY	541	tctgaatgagatgtgtgcacgitttgatgataagaacaatccttcccttataccaanaattgc	600
Db	1529	tctgaatgagatgtgtgcacgitttgatgataagaacaatccttcccttataccaanaattgc	1588
OY	601	tcaagtctcagaagaagcatctctaaanaacttggatcatatcatctgtctgtgaagaagagac	660
Db	1589	tcaagtctcagaagaagcatctctaaanaacttggatcatatcatctgtctgtgaagaagagac	1648
OY	661	tgggaactatgtctccctttagtctctgcgcccccgatgaagaag 701	
Db	1649	tgggaactatgtctccctttagtctctgcgcccccgatgaagaag 1689	

RESULT 4  
US-08-882-083-1  
; Sequence 1, Application US/08882083  
; Patent No. 5869292  
; GENERAL INFORMATION:  
; APPLICANT: VOOBERG, Johannes J

TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/882,083  
 FILING DATE:  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/558,107  
 FILING DATE: 13-NOV-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: ISACSON, John P.  
 REGISTRATION NUMBER: 33,715  
 REFERENCE/DOCKET NUMBER: 30472/212  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5035 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 35..5017  
 US-08-882-083-1

Query Match	Similarity	100.0%	Score	701	DB 2	Length	5035
Best Local	Similarity	100.0%	Pred. No.	4,76-213			
Matches	701	Conservative	0	Mismatches	0	Indels	0
							Gaps
							0
QY	1	atgtgagccctactagatgtagaagaaggagctctgcccagaagaaagacacagacctg	60				
Db	608	ATTGGAGCCCTACTAGTAGTGTAGAGAGGAGAGTCTGGCCAGAAAMGACACAGACTTG	667				
QY	61	cacaaattatatactacttlltgcgtgatttgatggaagggaaaagtgtggaaccagaanaa	120				
Db	668	CACAAATTTATCTACTCTTTTGGCTGATTGTGAAGAGGAAAAAGTTGGCACTAGAAACA	727				
QY	121	aagaactccttcatgacagatagagatagatgcatctgcctcgggacctgagcctaaatgca	180				
Db	728	AMGAACTCCTTATATGACGAGTAGGAGTAGTGCACTGTGCTGGGCGCTGGCCTAAATGAC	787				
QY	181	acagtcataatggtatgtaataacagagctctctgcgcaggtctgattgtagtgcacaggaata	240				
Db	788	ACAGTCAATAGGTTATGTAAACACAGTCTTGCCAGGTCTATTGGATGGCCACAGAAATCA	847				
QY	241	gtcatcttgcatatgtatgtgaatggtgcacacatccctggaatgtgcaatcaatattccctgaa	300				
Db	848	GTCATTATGGCATGTGATTGGAAATGGGACACACATCCCTGAAGTGCATCAATATTCCTCGAA	907				
QY	301	ggtcacacatcttctgtgtaggaacacatcgccagggtcccttggaatctcgccaataact	360				
Db	908	GGTCACACATTTCTGTGTGAGGAACCAATCGCCAGGCGCTCTGTGAAATCTCGCCAAATAACT	967				
QY	361	ttccttaacgtctcaaacacacttgataggaacctgtggacagtttactggtttgtcatatc	420				
Db	968	TTTCCTTAACGTCTCAACACACTCTTGATGGACCTTGGACAGTTTCTACTGTTTGTCAATAC	1027				

QY	421	ttctcccccacaacttgatggcatggaagctatgctcaagtagaagctgcttcagaagaa	480
Dp	1028	TTCTTCCACCACACATGATGGCATTGGAAGCTTATGTCAAAAGTACACGCTGTCCAAAGGAA	1087
QY	431	cccccaactgaatgtaaaaataatgaagaagcggaagactatgatgatcttactgat	540
Dp	1088	CCCCAACACACCAATGAAAAAATAATGAAAGACCGGAAGACTATGATGATGATCTTACTGAT	1147
QY	541	tctgaatatggaatggtgctcaggttctgatatgataaactcccttccttaccaaattgc	600
Dp	1148	TCTGAATGATGTGTGTCACAGTTTGTGATGATGACAACTCTCTTCTTATGCCAAATTTGGC	1207
QY	601	tccagtgcacaagaagactcctaaacttggatcattcaatctgctgcttgaagaagagac	660
Dp	1208	TCAAGTTCACAAAGAGCATCTCTAAACCTTGGGTACATTACATGTGTCGTAAGAGAGAC	1267
QY	661	tggagactatgctcccttaagtccttcgccccgcagatagaagaag	701
Dp	1268	TGGGACTATGTCTCTTAACTTATGTCCTGCGCCCGATGACAGAG	1308

```

1      RESULT          5
2      US-08-558-107-1
3      ; Sequence 1, Application US/08558107
4      ; Patent No. 5910481
5      ;
6      ; GENERAL INFORMATION:
7      ; APPLICANT: VOOERBERG, Johannes J.
8      ; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
9      ; NUMBER OF SEQUENCES: 17
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Foley & Lardner
12     ; STREET: 3000 K Street, N.W., Suite 500
13     ; CITY: Washington
14     ; STATE: D.C.
15     ; COUNTRY: USA
16     ; ZIP: 20007-5109
17     ;
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
23     ;
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/558, 107
26     ; FILING DATE: 13-NOV-1995
27     ; CLASSIFICATION: 435
28     ;
29     ; ATTORNEY/AGENT INFORMATION:
30     ; NAME: ISACSON, John P.
31     ; REGISTRATION NUMBER: 33,715
32     ; REFERENCE/DOCKET NUMBER: 30472/212
33     ;
34     ; TELECOMMUNICATION INFORMATION:
35     ; TELEPHONE: (202)672-5300
36     ; TELEFAX: (202)672-5399
37     ; TELE: 904136
38     ;
39     ; INFORMATION FOR SEQ ID NO: 1:
40     ; SEQUENCE CHARACTERISTICS:
41     ; LENGTH: 5035 base pairs
42     ; TYPE: nucleic acid
43     ; STRANDEDNESS: single
44     ; TOPOLOGY: linear
45     ;
46     ; FEATURE:
47     ; NAME/KEY: CDS
48     ; LOCATION: 35..5017
49     ;
50     ; US-08-558-107-1

```

Query Match	100.0%	Score 701	DB 2	Length 5035
Best Local Similarity	100.0%	Pred. No. 4,7e-213		
Matches 701	Conservative 0	Mismatches 0	Indels 0	Gaps 0

OY	1	atcgagagccctactagatgctagagaaagagctctgycgaagaaacacacagacctg	60
Db	608	ATTGAGACCCCTACTGTATGTAGACAAAGGAGGTCTGGCCAAAGGAAAGACACGACCTTG	667

QY	61	cccaaatlatacctcttgcgcatltaataagggaaagtctggacctaagaaca	120
Db	668	cacaatttatnacttcttggctgctatttgatagaaggaattggcactagaanaa	727
QY	121	aagaactccctgaatgcaggaataggaatgcctgcctcggccctgaatatgcac	180
Db	728	agaactcctcttgatgacagatagggatgacgtcactctcgccgcttgataatgacac	787
QY	181	acagacaaatggttatgtaaaagaatctctgcagagctctgatttgatgccaaggaata	240
Db	788	acagtcattggttatgtaaacaggtctctgcacagctctgatttgatgccaaggaata	847
QY	241	gtcatttggcatgtgatgtgaatlyggacacactcctcgaatgcacactaatlccctgaa	300
Db	848	gtctatttggcatgtgatgtgaatlyggacacacactcctcgaatgcacactaatlccctgaa	907
QY	301	ggtcaacatacttctgtgaggaacacatcgcagagcgctccttgaatactcgcgaataact	360
Db	908	ggtcaacatacttctgtgaggaacacatcgcagagcgctccttgaatactcgcgaataact	967
QY	361	ttccttactgtctcaaacactctgatgtgaccttggacagtttactgtttgtgcatatc	420
Db	968	tttccttactgtctcaaacactctgatgtgaccttggacagtttactgtttgtgcatatc	1027
QY	421	tccttccacaacatgaatggcatgaagactctatgltcaaatgataagcagcgtctccagaaggaa	480
Db	1028	tccttccacaacatgaatggcatgaagactctatgltcaaatgataagcagcgtctccagaaggaa	1087
QY	481	ccccacactacgaatgtaaaaaataatgaaagagcggaagactatgaatgaatcttactgat	540
Db	1088	ccccacactacgaatgtaaaaaataatgaaagagcggaagactatgaatgaatcttactgat	1147
QY	541	tcctgaatgagatgtgtgcaggttgaatgatgaacactcctctcccttcaataatgtgc	600
Db	1148	tcctgaatgagatgtgtgcaggttgaatgatgaacactcctctcccttcaataatgtgc	1207
QY	601	tcgattgtcccaagaagaactcctaaacttgggtacattatcatctgtcgtcgaagaggaagac	660
Db	1208	tcgattgtcccaagaagaagaactcctaaacttgggtacattatcatctgtcgtcgaagaggaagac	1267
QY	661	ttggagacatgtccctctagctcctcgcgcgcgcgaatgaaagaaag 701	
Db	1268	ttggagacatgtccctctagctcctcgcgcgcgcgaatgaaagaaag 1308	

RESULT 6  
 US-09-243-539-1  
 Sequence 1, Application US/09243359  
 Patent No. 6130203  
 GENERAL INFORMATION:  
 APPLICANT: VOORBERG, Johannes J.  
 TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/243,539  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/558,107  
 FILING DATE: 13-NOV-1995  
 ATTORNEY/AGENT INFORMATION:

NAME: ISACSON, John P.  
REGISTRATION NUMBER: 33,715  
REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5035 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 35..5017  
US-09-243-539-1

Query Match 100.0%; Score 701; DB 3; Length 5035;  
Best Local Similarity 100.0%; Pred. No. 4.7e-213;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgagccctactagtatgtatgagaagaggtctgcccagaagaaacacacagacttg 60  
DB attgagccctactagtatgtatgagaagaggtctgcccagaagaaacacacagacttg 667  
QY 61 cacaattatctactcttctgttattgaagaagaaagttggaactcagaaca 120  
DB cacaattatctactcttctgttattgaagaagaaagttggaactcagaaca 727  
QY 121 aagaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 180  
DB aagaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 787  
QY 728 aagaaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 787  
DB aagaaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 787  
QY 181 acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 240  
DB acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 847  
QY 788 acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 847  
DB acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 847  
QY 241 gtcatttgatgagatgagatgagacacactcctgaagtcaactcaatctcctgaa 300  
DB gtcatttgatgagatgagatgagacacactcctgaagtcaactcaatctcctgaa 907  
QY 848 gtcatttgatgagatgagatgagacacactcctgaagtcaactcaatctcctgaa 907  
DB gtcatttgatgagatgagatgagacacactcctgaagtcaactcaatctcctgaa 907  
QY 301 ggcacaactctctgtgagagacacacacacacacacacacacacacacacacacac 360  
DB ggcacaactctctgtgagagacacacacacacacacacacacacacacacacacac 967  
QY 908 ggcacaactctctgtgagagacacacacacacacacacacacacacacacacacac 967  
DB ggcacaactctctgtgagagacacacacacacacacacacacacacacacacacac 1027  
QY 361 ttcttactgctcaaacactctgagacacttgagacagtttctactgtttgtcatatc 420  
DB ttcttactgctcaaacactctgagacacttgagacagtttctactgtttgtcatatc 1027  
QY 968 ttcttactgctcaaacactctgagacacttgagacagtttctactgtttgtcatatc 1027  
DB ttcttactgctcaaacactctgagacacttgagacagtttctactgtttgtcatatc 1027  
QY 421 tcttccacaacaatgagatgagatgagatgagatgagatgagatgagatgagatgag 480  
DB tcttccacaacaatgagatgagatgagatgagatgagatgagatgagatgagatgag 1087  
QY 1028 tcttccacaacaatgagatgagatgagatgagatgagatgagatgagatgagatgag 1087  
DB tcttccacaacaatgagatgagatgagatgagatgagatgagatgagatgagatgag 1147  
QY 481 ccccaactcgaatgagatgagatgagatgagatgagatgagatgagatgagatgag 540  
DB ccccaactcgaatgagatgagatgagatgagatgagatgagatgagatgagatgag 1147  
QY 1088 ccccaactcgaatgagatgagatgagatgagatgagatgagatgagatgagatgag 1147  
DB ccccaactcgaatgagatgagatgagatgagatgagatgagatgagatgagatgag 1207  
QY 541 tctgaatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 600  
DB tctgaatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 1207  
QY 601 tctgaatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 660  
DB tctgaatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 1267  
QY 661 tggagactatgctcctctgctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 701  
DB tggagactatgctcctctgctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1308  
QY 1268 tggagactatgctcctctgctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1308  
DB tggagactatgctcctctgctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1308

RESULT 7

US-08-276-594A-1  
Sequence 1, Application US/08276594A  
Patent No. 5693499

GENERAL INFORMATION:  
APPLICANT: YONEMURA, Hiroshi  
APPLICANT: TAJIMA, Yoshitaka  
APPLICANT: SUGAMURA, Keishin  
APPLICANT: MASUDA, Kenichi  
TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION  
TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/276,594A  
FILING DATE: 18-JUL-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/950,191  
FILING DATE: 24-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 243262/1991  
FILING DATE: 24-SEP-1991

ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/195/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6999 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6996  
US-08-276-594A-1

Query Match 100.0%; Score 701; DB 1; Length 6999;  
Best Local Similarity 100.0%; Pred. No. 5.6e-213;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgagccctactagtatgtatgagaagaggtctgcccagaagaaacacacagacttg 60  
DB attgagccctactagtatgtatgagaagaggtctgcccagaagaaacacacagacttg 576  
QY 517 attgagccctactagtatgtatgagaagaggtctgcccagaagaaacacacagacttg 576  
DB attgagccctactagtatgtatgagaagaggtctgcccagaagaaacacacagacttg 636  
QY 61 cacaattatctactcttctgttattgaagaagaaagttggaactcagaaca 120  
DB cacaattatctactcttctgttattgaagaagaaagttggaactcagaaca 636  
QY 121 aagaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 180  
DB aagaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 696  
QY 637 aagaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 696  
DB aagaactccttgatgagagatgagatgctgctcggcctgagccttaataatgac 756  
QY 181 acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 240  
DB acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 756  
QY 697 acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 756  
DB acagtaaatggtatgtaaaacaggtctcgcaggtctgattgagatgacacagaaatca 756

```

QY 241 gtctattgcatgattgaaatggcaccactcctgaagtcactcaatctctcgaa 300
    |||||||
Db 757 gtctattgcatgattgaaatggcaccactcctgaagtcactcaatctctcgaa 816
    |||||||
QY 301 ggtcaacacttctgttgagggaacatcgccagcgctcttggaatctcgcaataact 360
    |||||||
Db 817 ggtcaacacttctgttgagggaacatcgccagcgctcttggaatctcgcaataact 876
    |||||||
QY 361 ttccttaagtcacacactctgtgacacttgacagtttctctcttctgtatc 420
    |||||||
Db 877 ttccttaagtcacacactctgtgacacttgacagtttctctcttctgtatc 936
    |||||||
QY 421 tcttccaccacatgattgacatggaaagcttatgtcaaaagtagacagctgtccagaagaa 480
    |||||||
Db 937 tcttccaccacatgattgacatggaaagcttatgtcaaaagtagacagctgtccagaagaa 996
    |||||||
QY 481 ccccaactagatgataaataatgaagaagcggaaagactatgatgatcttactgat 540
    |||||||
Db 997 ccccaactagatgataaataatgaagaagcggaaagactatgatgatcttactgat 1056
    |||||||
QY 541 tctgaatgagatgtgcaagtttgatgatgacaactctccttcttatacgaattcgc 600
    |||||||
Db 1057 tctgaatgagatgtgcaagtttgatgatgacaactctccttcttatacgaattcgc 1116
    |||||||
QY 601 tcaagttgccaaagaagcatccttaaaacttggtgaacttaacttgctgctgaagaagagac 660
    |||||||
Db 1117 tcaagttgccaaagaagcatccttaaaacttggtgaacttaacttgctgctgaagaagagac 1176
    |||||||
QY 661 tgggaactagtccttctgctcctcgcggcgatgacgaag 701
    |||||||
Db 1177 tgggaactagtccttctgctcctcgcggcgatgacgaag 1217
    |||||||

```

# RESULT 8

US-08-121-202-1  
Sequence 1, Application US/08121202  
Patent No. 5563045

## GENERAL INFORMATION:

APPLICANT: Pitman, Debra  
APPLICANT: Rehmetulla, Alnawaz  
APPLICANT: Wozney, John M.  
APPLICANT: Kaufman, Randal J.  
TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/121, 202  
FILING DATE: 14-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meinert, M. C.  
REGISTRATION NUMBER: 31,544  
REFERENCE/DOCKET NUMBER: GI 5195A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1210 X8574  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7056 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7053  
US-08-121-202-1

Query Match 100.0%; Score 701; DB 1; Length 7056;  
Best Local Similarity 100.0%; Pred. No. 5,6e-213;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 attggaacccctactagatgtaagaaggagctctggccaaggaagaagacagacttg 60
    |||||||
Db 574 attggaacccctactagatgtaagaaggagctctggccaaggaagaagacagacttg 633
    |||||||
QY 61 cacaacttactactcttctgtatattgtaagaagggaagattggcactcagaagaa 120
    |||||||
Db 634 cacaacttactactcttctgtatattgtaagaagggaagattggcactcagaagaa 693
    |||||||
QY 121 aagaactccttgatgcaaggaatgcatctgctcctgagcctgagcctaaatgcaac 180
    |||||||
Db 694 aagaactccttgatgcaaggaatgcatctgctcctgagcctgagcctaaatgcaac 753
    |||||||
QY 181 acagtcactgattatgaacagctctctgcaagctctgattggaatgacacaggaatca 240
    |||||||
Db 754 acagtcactgattatgaacagctctctgcaagctctgattggaatgacacaggaatca 813
    |||||||
QY 241 gtctattgcatgattgaaatggcaccactcctgaagtcactcaatctctcgaa 300
    |||||||
Db 814 gtctattgcatgattgaaatggcaccactcctgaagtcactcaatctctcgaa 873
    |||||||
QY 301 ggtcaacacttctgttgagggaacatcgccagcgctcttggaatctcgcaataact 360
    |||||||
Db 874 ggtcaacacttctgttgagggaacatcgccagcgctcttggaatctcgcaataact 933
    |||||||
QY 361 ttccttaagtcacacactctgtgacacttgacagtttctctcttctgtatc 420
    |||||||
Db 934 ttccttaagtcacacactctgtgacacttgacagtttctctcttctgtatc 993
    |||||||
QY 421 tcttccaccacatgattgacatggaaagcttatgtcaaaagtagacagctgtccagaagaa 480
    |||||||
Db 994 tcttccaccacatgattgacatggaaagcttatgtcaaaagtagacagctgtccagaagaa 1053
    |||||||
QY 481 ccccaactagatgataaataatgaagaagcggaaagactatgatgatcttactgat 540
    |||||||
Db 1054 ccccaactagatgataaataatgaagaagcggaaagactatgatgatcttactgat 1113
    |||||||
QY 541 tctgaatgagatgtgcaagtttgatgatgacaactctccttcttatacgaattcgc 600
    |||||||
Db 1114 tctgaatgagatgtgcaagtttgatgatgacaactctccttcttatacgaattcgc 1173
    |||||||
QY 601 tcaagttgccaaagaagcatccttaaaacttggtgaacttaacttgctgctgaagaagagac 660
    |||||||
Db 1174 tcaagttgccaaagaagcatccttaaaacttggtgaacttaacttgctgctgaagaagagac 1233
    |||||||
QY 661 tgggaactagtccttctgctcctcgcggcgatgacgaag 701
    |||||||
Db 1234 tgggaactagtccttctgctcctcgcggcgatgacgaag 1274
    |||||||

```

# RESULT 9

US-08-121-202-1  
Patent No. 5171844

APPLICANT: VAN OYEN, ALBERT J.J.; PANNEKOEK, HANS; VERBEET, MARTINUS P.; VAN LEEN, ROBERT W.  
TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY  
PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS  
AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM  
NUMBER OF SEQUENCES: 12  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/205, 226

FILING DATE: 10-JUN-1988

SEQ ID NO:1

LENGTH: 8241

5171844-1

Query Match 100.0%; Score 701; DB 6; Length 8241;  
Best Local Similarity 100.0%; Pred. No. 6.1e-213;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 attgagccctactagatgtatgaagaagagatcgtgccaagaaagacacagacttg 60  
|||||  
580 attgagccctactagatgtatgaagaagagatcgtgccaagaaagacacagacttg 639  
|||||  
61 cacaattatactacttttctgtatattgaagaagaaagttgacctgaagaaca 120  
|||||  
640 cacaattatactacttttctgtatattgaagaagaaagttgacctgaagaaca 699  
|||||  
121 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 180  
|||||  
700 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 759  
|||||  
181 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 240  
|||||  
760 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 819  
|||||  
241 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 300  
|||||  
820 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 879  
|||||  
301 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 360  
|||||  
880 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 939  
|||||  
361 ttcctactgctcaacaactcttgatggacacttgacagtttctactgtttgtcatac 420  
|||||  
940 ttcctactgctcaacaactcttgatggacacttgacagtttctactgtttgtcatac 999  
|||||  
421 tcttcccaacaatgataggcagatgataatgataatgataatgataatgataatgata 480  
|||||  
1000 tcttcccaacaatgataggcagatgataatgataatgataatgataatgataatgata 1059  
|||||  
481 cccaactcgaatgaataaataatgaagaagcgaagactgacatgacatgacatgacatgac 540  
|||||  
1060 cccaactcgaatgaataaataatgaagaagcgaagactgacatgacatgacatgacatgac 1119  
|||||  
541 tctgaatgcatgtgctgacgtttgatagaacaactcctctcttataccaattcgc 600  
|||||  
1120 tctgaatgcatgtgctgacgtttgatagaacaactcctctcttataccaattcgc 1179  
|||||  
601 tcaagtgcagaagaagcactcctaacttgggtacattatgctgctgataagaagagac 660  
|||||  
1180 tcaagtgcagaagaagcactcctaacttgggtacattatgctgctgataagaagagac 1239  
|||||  
661 tggagactgctcctctgacgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 701  
|||||  
1240 tggagactgctcctctgacgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1280  
|||||

RESULT 10  
US-08-366-851A-1  
Sequence 1, Application US/08366851A

Patent No. 5681746  
GENERAL INFORMATION:

APPLICANT: Bodner, Mordechai

APPLICANT: De Polo, Nicolas J.

APPLICANT: Hsu, David Chi-Tang

APPLICANT: Chang, Steven

TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Viagene, Inc.

STREET: 11055 Roselle Street

CITY: San Diego  
STATE: California  
COUNTRY: U.S.A.

ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/366,851A

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Chambers, Daniel M.

REGISTRATION NUMBER: 34,561

REFERENCE/DOCKET NUMBER: 930049, 438

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 452-1288

TELEFAX: (619) 452-2616

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8967 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: unknown

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 110..7165

US-08-366-851A-1

Query Match 100.0%; Score 701; DB 1; Length 8967;  
Best Local Similarity 100.0%; Pred. No. 6.4e-213;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 attgagccctactagatgtatgaagaagagatcgtgccaagaaagacacagacttg 60  
|||||  
683 attgagccctactagatgtatgaagaagagatcgtgccaagaaagacacagacttg 742  
|||||  
61 cacaattatactacttttctgtatattgaagaagaaagttgacctgaagaaca 120  
|||||  
743 cacaattatactacttttctgtatattgaagaagaaagttgacctgaagaaca 802  
|||||  
121 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 180  
|||||  
803 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 862  
|||||  
181 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 240  
|||||  
863 aagaactcttgatgacagatagaagatgctgacatgctgctggcctgaagaatgac 922  
|||||  
241 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 300  
|||||  
923 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 982  
|||||  
301 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 360  
|||||  
983 gctatctgcatgtatgtgaatgggacacactctgaagtgcactcaatattctcgaa 1042  
|||||  
361 ttcctactgctcaacaactcttgatggacacttgacagtttctactgtttgtcatac 420  
|||||  
1043 ttcctactgctcaacaactcttgatggacacttgacagtttctactgtttgtcatac 1102  
|||||  
421 tcttcccaacaatgataggcagatgataatgataatgataatgataatgataatgata 480  
|||||  
1103 tcttcccaacaatgataggcagatgataatgataatgataatgataatgataatgata 1162  
|||||  
481 cccaactcgaatgaataaataatgaagaagcgaagactgacatgacatgacatgacatgac 540  
|||||  
1163 cccaactcgaatgaataaataatgaagaagcgaagactgacatgacatgacatgacatgac 1222  
|||||



QY 541 tcgaatggatgtgtgtgcgaagtttgatgtgaaacatcccttccttataccaatttcgc 600

Db 1223 TCTAAATGGATGTGTGGCAAGTTTGATGTGATGACAACCTCCCTTATCCAAATTTGCC 1282

QY 601 tcagttccagaagaacatccataacttggatcatcatctgtcgcgcgaagaagagagac 660

Db 1283 TCAATTCGCAAGAGCATCTCTAAACTTTGGGTACATTTACATTGTGCTGTAAGAGGAGAGAC 1342

QY 661 ttggactatgtcccttaagtctcctcgcgcccgatgacaagaag 701

Db 1343 TGGGACTATGCTCCCTTAAGTCCTCGGCCGCCGATGACAGAG 1383

```

RESULT 11
US-07-864-004B-3
; Sequence 3, Application US/07864004B
; Patent No. 5364771
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/864,004B
; FILING DATE: 07 APRIL 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patricia L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9009 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver
; FEATURE:
; NAME/KEY: misc_feature (Domain Structure)
; LOCATION: 5001 . . . 7053
; OTHER INFORMATION: /note= "Equivalent to the A3-C1-C2
; OTHER INFORMATION: domain"
; FEATURE:
; NAME/KEY: misc_feature (Domain Structure)
; LOCATION: 1 . . . 2277
; OTHER INFORMATION: /note= "Equivalent to the A1-A2
; OTHER INFORMATION: domain"
US-07-864-004B-3
Query Match 100.0%; Score 701; DB 1; Length 9009;
Best Local Similarity 100.0%; Pred. No. 6,4e-213;

```

Matches	701:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
QY	1	attggagccctactagtatgtagaagaaggagtcgycgcaaggaagaaacacagacctg	60						
Db	724	ATTGGAGCCCTACTAGTATGTAGAGAAAGGAGTCTGGCCAAAGAAAAGACAGACCTTG	783						
QY	61	cacaaatttactactcttcttgcgtatcttgatgaagggaaaagttggnactcaagaaca	120						
Db	784	CACAAATTTTACTACTCTTTTGGCTGTATTTGATGAAGGGAAGAAAGTGGACCTCAGAAATA	843						
QY	121	aagaacctctgaatgcagaataaggatgtctgcatactgcgcgagccctgacctaaatgcac	180						
Db	844	AAGAATCCTTGATGCAGAGTAGGAGTGGTGCATCTGCTCGGGCCCTGGCCTAAATGCAAC	903						
QY	181	acagtcataatggtatgtataaacsaggtctctgcaggtctgatatggaatgcacagaanaatca	240						
Db	904	ACAGTCAAATGGTTATGTAAACAGAGTCTGCACAGCTCTGATTTGGATGCCACAGGAATTA	963						
QY	241	gtctatctggcatgtatgttggaaatggaccactctcgtgaatgtcacctcaatattctctgaa	300						
Db	964	GTCATATGGCATGTGATTTGGAATGGGACACCACTCTCGAAGTGCACCAATATTCCTCGAA	1023						
QY	301	ggtcacacattctctgttgaagacacatgcacagcgctctcttgaaatcctgcacaaataact	360						
Db	1024	GGTCACACATTTCTTGTGTGAGGAACCATGCCACGGCGTCTTGGAAATCTGCCAATAACT	1083						
QY	361	tctccttactgcgtcaaacactcttgatbgaacctttgacagttcttactctgtttgtgcatactc	420						
Db	1084	TTCTCTTACTGCTCAAAACACTCTTTGATGGACCTTGGACACTTTCTCATGTTTGTCATATTC	1143						
QY	421	tcttcaccacacatgatatgagcatggaaagcttctgcacaaagtatgaacgtcttcacaaagaa	480						
Db	1144	TTTTCOCACCAACATGATGTGCAATGATTAATGTCAAAGTGAAGACGCTGTCCACAGGAA	1203						
QY	481	cccccaactacgaatgaaaaataatataaagaagcggagaagactatgcatgcatcttaactgat	540						
Db	1204	CCCCAAGCTTTCGCAATGAAAAATATATGAAGAAGGGGAAGACTATGATGATGATTTACTGAT	1263						
QY	541	tctgaaatgagatgtgtgcaggtctgatatgcatacgaactcctcctcttataccaaatctgcg	600						
Db	1264	TCTGAAATGGAGTGTGTGAGGTTTGTATGATGACCAACTCTCTTATTCCAAATTTGCG	1323						
QY	601	tcaagttgcacaaagaacatcctaaacacttgggtacacttcatcttgcctcgtgaagaagaagac	660						
Db	1324	TCAAGTTGCCAAGAACATCTTAAAACTTGGGTGACATTTGCTGCTGAAGAGAGAGGAC	1383						
QY	661	tggagactatgctcccttagtctctgcgcccccgatgcagaag	701						
Db	1384	TGGAGACTATGCTCCCTTAGTCTCTGCCCCCGATGACAGAAAG	1424						

RESULT 12  
US-08-251-937A-3  
Sequence 3, Application US/08251937A  
Patent No. 5583209  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.  
APPLICANT: Runge, Merschal S.  
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:







QY 301 ggtcacacattctgtgaggaacacatcgccaggcgltccttggaatctcgcaataact 360  
DB 1024 ggtcacacattctgtgaggaacacatcgccaggcgltccttggaatctcgcaataact 1083  
QY 361 ttcttactgtctcaaacactctgtatggaacttgaacagtttctactgtttgtcatatc 420  
DB 1084 ttcttactgtctcaaacactctgtatggaacttgaacagtttctactgtttgtcatatc 1143  
QY 421 tcttcccccacacatgtaggcatggaactatgtcaagtagagacactgtccagagaa 480  
DB 1144 tcttcccccacacatgtaggcatggaactatgtcaagtagagacactgtccagagaa 1203  
QY 481 ccccaactacgaatgaaataatgaagaacggaagactatgtagatctactctgat 540  
DB 1204 ccccaactacgaatgaaataatgaagaacggaagactatgtagatctactctgat 1263  
QY 541 tctgaatgtagtgtgtcaggttgaatgtagaactctcttcttccatccaaatcgc 600  
DB 1264 tctgaatgtagtgtgtcaggttgaatgtagaactctcttcttccatccaaatcgc 1323  
QY 601 tcaattgccaagaagacatccaaacttggtatcattacattgtctgtgaagagagac 660  
DB 1324 tcaattgccaagaagacatccaaacttggtatcattacattgtctgtgaagagagac 1383  
QY 661 tgggaactatgctcccttagtctcgcgccccgatgacagaag 701  
DB 1384 tgggaactatgctcccttagtctcgcgccccgatgacagaag 1424

RESULT 15  
US-08-670-707A-1  
Sequence 1, Application US/08670707A  
Patent No. 5859204

## GENERAL INFORMATION:

APPLICANT: Lollar, John S.  
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/670,707A  
FILING DATE: 26-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US94/13200  
FILING DATE: 15-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,133  
FILING DATE: 11-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,004  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenlee, Lorraine L.  
REGISTRATION NUMBER: 27,894  
REFERENCE/DOCKET NUMBER: 75-95F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8089  
TELEFAX: 303/499-8089

SEQUENCE CHARACTERISTICS:  
SEQUENCE FOR SEQ ID NO: 1:  
LENGTH: 9009 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Liver  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..2277  
OTHER INFORMATION: /product= "Domain Structure"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..2277  
OTHER INFORMATION: /note= "Equivalent to the A1-A2 domain"  
OTHER INFORMATION: /product= "Domain"  
OTHER INFORMATION: /note= "cDNA encoding human factorVIII"  
US-08-670-707A-1

Query Match 100.0%; Score 701; DB 2; Length 9009;  
Best Local Similarity 100.0%; Fred. No. 6.4e-213;  
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attggaagccctactatgtagaagaagagatctgccaagaagaacacagacctg 60  
DB 724 attggaagccctactatgtagaagaagagatctgccaagaagaacacagacctg 783  
QY 61 cacaattatctactcttctgtatctgtagaaggaagaagtgtgactcagaaca 120  
DB 784 CACAATTATCTACTCTTCTGTATTTGATGAGGAGGAAAAAGTGGCAGTCAGAAACA 843  
QY 121 aaaaactccttgaatgaagaatgaagatgctgcatctgcgcggcgctgaataatgac 180  
DB 844 AAAAACTCCTTGATGCAAGATGAGATGCTGCTGCTGGGCTTGCTTAAATGACAC 903  
QY 181 acagtaatgtagtataaacaaggtctgcgcagctgctgtagtgaatgcacagaaata 240  
DB 904 ACAGTAAATGATATGTAACACAGCTCTGCGCAGCTGATGAGACACAGGAATGA 963  
QY 241 gttcattgcatgtgtagaattgggacacactcctgaagtgaactcaatctcgtgaa 300  
DB 964 GTCTATTGGCATGTGATTGGAATGGGACACACCTCGAAGTGCACATATTCCTCGAA 1023  
QY 301 ggtcacacattctgtgaggaacacatcgccaggcgltccttggaatctcgcaataact 360  
DB 1024 ggtcacacattctgtgaggaacacatcgccaggcgltccttggaatctcgcaataact 1083  
QY 361 ttcttactgtctcaaacactctgtatggaacttgaacagtttctactgtttgtcatatc 420  
DB 1084 ttcttactgtctcaaacactctgtatggaacttgaacagtttctactgtttgtcatatc 1143  
QY 421 tcttcccccacacatgtaggcatggaactatgtcaagtagagacactgtccagagaa 480  
DB 1144 tcttcccccacacatgtaggcatggaactatgtcaagtagagacactgtccagagaa 1203  
QY 481 ccccaactacgaatgaaataatgaagaacggaagactatgtagatctactctgat 540  
DB 1204 ccccaactacgaatgaaataatgaagaacggaagactatgtagatctactctgat 1263  
QY 541 tctgaatgtagtgtgtcaggttgaatgtagaactctcttcttccatccaaatcgc 600  
DB 1264 tctgaatgtagtgtgtcaggttgaatgtagaactctcttcttccatccaaatcgc 1323  
QY 601 tcaattgccaagaagacatccaaacttggtatcattacattgtctgtgaagagagac 660

Db 1324 TCAGTTGCCAAGAGCATCCTAAACTTGGGTACATTACATTGCTGCTGAAGAGGAGAC 1383

Oy 661 tgggactatgctcccttagtctcgcccccgatgacagaag 701

Db 1384 TGGGACTATGCTCCCTTAGTCTCGCCCCCGATGACAGAAAG 1424

Search completed: June 13, 2002, 22:40:35  
Job time: 18083 sec



D	b		241	gtgcagtagtcgcccgtgaacgtctcttttgcgaacgggttcgcccccgccaggta	300
O	y		301	gtgccagggaatgttttgtcttaaatacacatcgtcccaagggaattgttcttaaatc	360
D	b		301	gtgccagggaatgttttgtcttaaatacacatcgtcccaagggaattgttcttaaatc	360
O	y		361	cactaacgaaacacgaaccacattcttctttctccacagatatgatccacaatga	420
D	b		361	cactactcgaacacgaaccacattcttctttctccacagatatgatccacaatga	420
O	y		421	aatagaagctccacccctgtctcttctctgtgccttttgcatctgcttagtagtcca	480
D	b		421	aatagaagctccacccctgtctcttctctgtgccttttgcatctgcttagtagtcca	480
O	y		481	aagaatacctaactgtgacgtgacagtcaataggactataatgaagatctgtgta	540
D	b		481	aagaatacctaactgtgacgtgacagtcaataggactataatgaagatctgtgta	540
O	y		541	gtctgcctgtgacgcgaagatctccctccaaagtagcacaatctctccattcaacac	600
D	b		541	gtctgcctgtgacgcgaagatctccctccaaagtagcacaatctctccattcaacac	600
O	y		601	agtcgtatcaaaaaaagacctcgtttgttagaattcaacggtaccctttcaacatgc	660
D	b		601	agtcgtatcaaaaaaagacctcgtttgttagaattcaacggtaccctttcaacatgc	660
O	y		661	gccaaagccaccctcgagatggtcgtcgaagtcctcacatccacgaagctgaagttat	720
D	b		661	gccaaagccaccctcgagatggtcgtcgaagtcctcacatccacgaagctgaagttat	720
O	y		721	agtggtcattaacacttaagaacatgagcttccacatccatcgtcagtcctcaatgc	780
D	b		721	agtggtcattaacacttaagaacatgagcttccacatccatcgtcagtcctcaatgc	780
O	y		781	atccactcgtgaagcttcctg	800
D	b		781	atccactcgtgaagcttcctg	800
RESULT 2					
US-09-364-862-14					
Sequence 14, Application US/09364862					
Patent No. 6221349					
GENERAL INFORMATION:					
APPLICANT: Coulo, Linda B.					
APPLICANT: Colosi, Peter C.					
TITLE OF INVENTION: ADENO-ASSOCIATED VECTORS FOR EXPRESSION OF FACTOR VIII					
TITLE OF INVENTION: BY TARGET					
FILE REFERENCE: AVIGEN-03743					
CURRENT APPLICATION NUMBER: US/09/364, 862					
CURRENT FILING DATE: 1999-07-30					
EARLIER APPLICATION NUMBER: 60/125, 974					
EARLIER FILING DATE: 1999-03-24					
EARLIER APPLICATION NUMBER: 60/104, 994					
EARLIER FILING DATE: 1998-10-20					
NUMBER OF SEQ ID NOS: 14					
SOFTWARE: PatentIn Ver. 2.0					
SEQ ID NO 14					
LENGTH: 4999					
TYPE: DNA					
ORGANISM: Artificial Sequence					
FEATURE:					
OTHER INFORMATION: Description of Artificial Sequence: Synthetic					
US-09-364-862-14					

Query Match	100.0%	Score 800;	DB 4;	Length 4999;
Best Local Similarity	100.0%;	Pred. No. 1.5e-241;		
Matches 800; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible]

```

RESULT      3
US-09-470-618-13
; Sequence 13, Application US/09470618
; Patent No. 6200560
; GENERAL INFORMATION:
; APPLICANT: Couto, Linda B.
; APPLICANT: Colosi, Peter C.
; TITLE OF INVENTION: Adeno-Associated Vectors for Expression of Factor VIII
; TITLE OF INVENTION: by Target Cells
; FILE REFERENCE: Avigen-04082
; CURRENT APPLICATION NUMBER: US/09/470.618
; CURRENT FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: 09/364,862
; EARLIER FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: 60/125,974
; EARLIER FILING DATE: 1999-03-24
; EARLIER APPLICATION NUMBER: 60/104,994

```





Db	601	ttcaaacactcagtcggttacaagaagactctggtttttagaatccaagctaacctttc	660
Oy	650	aacatcgctaaagccaagagccaccccttgatggtctgtgaagtcctaccatccaagctgag	709
Db	661	aacacgcgtaaagccaagagccaccccttgatggtctgtgaagtcctaccatccaagctgag	720
Oy	710	gttatatgaataaagtcgttcaatcaccttaagaacatgctctccactcctgtcaatcttcat	769
Db	721	gttatatgaataaagtcgttcaatcaccttaagaacatgctctccactcctgtcaatcttcat	780
Oy	770	gcctgttgctatctactactgtaaaagctcttg	800
Db	781	gcctgttgctatctactactgtaaaagctcttg	811

RESULT 5  
US-08-882-083-1

; Sequence 1, Application US/08882083  
; Patent No. 5869292  
; GENERAL INFORMATION:

```

; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
;

```

```

; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
;
; 00007 7100

```

```

; ZIP: 2000/-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy
; CONTAINED: TBM PC 300

```

```

;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/882,083
FILING DATE:
CLASSIFICATION:  51A

```

CLASSIFICATION: 214  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/558,107  
FILING DATE: 13-NOV-1995

ATTORNEY/AGENT INFORMATION:  
NAME: ISACSON, John P.  
REGISTRATION NUMBER: 33,715

REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300

TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 5035 base pairs
; TYPE: nucleic acid
;

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 35..5017
US-08-882-083-1

```

Query Match	49.0%	Score 392.2	DB 2	Length 5035
-------------	-------	-------------	------	-------------

Best Local Similarity 96.9%; Pred. No. 2.7e-113;  
Matches 400; Conservative 0; Mismatches 13; Indels 0; Gaps 0

QY 388 tcttttctccacagcgtatcgatccacacatgcaaatagagtctccacctgtctcttc 44  
|| || || || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 7 TCCAGTTGACATTGTATGCAAGCCACCATGGAAATAGAGCTCCACACTGCTTCTTTC 66

Qy 448 gtgccttttgcgattctgctttagtgccaccagaagatactactcttggtgcagtgaaat 507  
|||||

Db	67	GTGCTTTTGGATTCGTTGCTTTAGTCGACACCAAGAACTACTACTGCGTACAGTGAACT	126
QY	508	gtcatgagcatcatatgcaagaatgacatctcagttgtagctctccgttgagccaagatttccctcc	567
Db	127	GTCATGGGACTATATGCAAAAGTGATTCGGGAGACTCCCTGTGGACCAAGATTTCTCC	186
QY	568	tagaatgccaaacatttttccatccaacacctcagtcgtgacaaaagaatcctgtttgt	627
Db	187	TAGAGTGGCAAAATCTTTTCATTCACACCTCAGTCGTGTACAAAAGACTCGTTTGT	246
QY	628	agaattcaagataaccttttcaacatcgcgtlaagccaagccaacctgagatgctgcct	687
Db	247	AGAATTCAAGATTAACCTTTTCAACATCGCTAAGCCAAAGCCACCTCGATGGGTCTGCT	306
QY	688	agatctaccacatccagcgtcgaagttaatlgacagatggttcattacacctaagaacatgac	747
Db	307	AGGTCCTAACCATCCACAGCGCTGAGGTTTATGTACAGTGCCTATTAACTTAAGAAACATGGC	366
QY	748	ttcccatccctgacagcttcaatgctgtctgttgatctccctacatggaagacctctg	800
Db	367	TTCCCATCTCTCAAGTCTTCACTGCTGTGGGTATCTCTACTGCAAAACTTCTG	419

## RESULT 6

US-08-558-107-1  
; Sequence 1, Application US/08558107  
; Patent No. 5910481  
ATTORNEY: J. J. HARRINGTON

```

; GENERAL INFORMATION:
;
; APPLICANT: VOORBERG, Johannes J.
;
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
;
; NUMBER OF SEQUENCES: 17

```

NUMBER OF SEQUENCES: 1/  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
CMBREM: 3000 v Stroot N W Suite 500

STATE: D.C.  
CITY: Washington  
SIREN: 5000 N  
COUNTRY: USA

COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY

```
;
; MEDIUM FILE. COPY CERN
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS
; SOFTWARE: PatentIn Release
```

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/5558,107  
 FILING DATE: 13-NOV-1995

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ISACSON, John P.

REGISTRATION NUMBER: 33,715  
REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136

```
; INFORMATION FOR SEQ ID NO: 1
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5035 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;

```

```

;          FEATURE:
;          NAME/KEY: CDS
;          LOCATION: 35..5017
;          00 570 107 1

```

OS-08-558-107-1

Query Match	49.08;	Score 392.4;	DB 2	Length 5033;
Best Local Similarity	96.98;	Pred. No. 2.7e-113;		
Matches 400; Conservative	0;	Mismatches 13;	Indels 0;	Gaps 0;

Qy 388 tcttttccacagtatcgatccaccatgcgaatatagagctcctcacctgtcttctc 447  
|| || | ||| | | ||||||| ||||||| ||||||| |||||||





1 APPLICANT: Runge, Marschall S.  
 2 TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
 3 NUMBER OF SEQUENCES: 10  
 4 CORRESPONDENCE ADDRESS:  
 5 ADDRESSEE: Kilpatrick & Cody  
 6 STREET: 1100 Peachtree Street  
 7 City: Atlanta  
 8 STATE: Georgia  
 9 COUNTRY: US  
 10 ZIP: 30309  
 11  
 12 COMPUTER READABLE FORM:  
 13 MEDIUM TYPE: Floppy disk  
 14 COMPUTER: IBM PC compatible  
 15 OPERATING SYSTEM: PC-DOS/MS-DOS  
 16 SOFTWARE: Patent Release #1.0, Version #1.25  
 17  
 18 CURRENT APPLICATION DATA:  
 19 APPLICATION NUMBER: US/08/251,937A  
 20 FILING DATE: 31-MAY-1994  
 21 CLASSIFICATION: 435  
 22 PRIORITY APPLICATION DATA:  
 23 APPLICATION NUMBER: US 07/864,004  
 24 FILING DATE: 07-APR-1992  
 25 ATTORNEY/AGENT INFORMATION:  
 26 NAME: Pratt, John S.  
 27 REGISTRATION NUMBER: 29,476  
 28 REFERENCE/DOCKET NUMBER: EMU106DIV  
 29 TELECOMMUNICATION INFORMATION:  
 30 TELEPHONE: 404-815-6367  
 31 TELEFAX: 404-815-6555  
 32  
 33 INFORMATION FOR SEQ ID NO: 3:  
 34 SEQUENCE CHARACTERISTICS:  
 35 LENGTH: 9009 base pairs  
 36 TYPE: nucleic acid  
 37 STRANDEDNESS: single  
 38 TOPOLOGY: linear  
 39 MOLECULE TYPE: cDNA  
 40 HYPOTHETICAL: NO  
 41 ANTI-SENSE: NO  
 42 FRAGMENT TYPE: N-terminal  
 43 ORIGINAL SOURCE:  
 44 ORGANISM: Homo sapien  
 45 TISSUE TYPE: Liver  
 46  
 47 FEATURE:  
 48 NAME/KEY: misc\_feature (Domain Structure)  
 49 LOCATION: 5001...7053  
 50 OTHER INFORMATION: /note="Equivalent to the A3-C1-C2  
 51 OTHER INFORMATION: domain"  
 52  
 53 FEATURE:  
 54 NAME/KEY: misc\_feature (Domain Structure)  
 55 LOCATION: 1...2277  
 56 OTHER INFORMATION: /note="Equivalent to the A1-A2  
 57 OTHER INFORMATION: domain"  
 58  
 59  
 60

Query	March	Similarity	48.6%	Score	389	DB 1	Length	9009	
Best	Local	Similarity	94.2%	Pred.	No. 3.7e-112				
Matches	404	Conservative	0	Mismatches	22	Indels	0	Gaps	0

OY	372	actgaacatccacctttcttcttccacagatcgcacccatgcaaatagagctct	431
Db	107	ATTAACTTTTGCTCTCCAGTTGAACATTTTGTACCAATAATGTCAATATGAGACTCT	166
OY	432	ccacctgctctctcttgctgacctttgacatctgctttagtgcacacgaagatactacc	491
Db	167	CCACCTGCTTTCTTTCGTCGTCCTTTGGCGATTTGTGTCGCCACCAAAATATCTACC	226
OY	492	tgggtgcagtggaactgctatgtaggactatataccaagtatctcggtgagccgcctgtg	551
Db	227	tgggtgcacatggaactgctatgtaggactatataccaagtatctcggtgagccgcctgtg	286
OY	552	acgcaagaattctctctagatgccaacactctttccatccaacacctgaftcgtgtaaa	611

Db	287	ACGCAAGTTTCCTCCCTAGATGCGCAAAATCTTTTCATTCACACCTGAGTGGTAC	346
Qy	612	aaaagactctgttttgtagaatcaacgatacctttcaacatcgcataagccaagggccc	671
Db	347	AAAAGACCTCTTTGTGAAATTCACGGTTCACCTTTTCAACATGCTTAAGCCAAAGGCC	406
Qy	672	cctgataggctcctctagctccataccatccagcgtgaggtlatacagtggtcat	731
Db	407	CCTGATGGGCTCTGCTAGGCTCTTACATCCAGGCTGAGGTTATGATACAGTGGTCTATTA	466
Qy	732	cacttaagaacatggtcttccatccctctcagttctctatgcttggtgtagtactactcaga	791
Db	467	CACTTAAACAATGGCTTCCATCTGTCACTTCATGCTGTGGTGTATCTTACTGGA	526
Qy	792	aagctctcgc 800	
Db	527	AAGCTTCTG 535	

```

1  RESULT 11
2  US-08-212-133A-1
3  ? Sequence 1, Application US/08212133A
4  ? Patent No. 5663060
5  ? GENERAL INFORMATION:
6  ? APPLICANT: Lollar, John S.
7  ? APPLICANT: Runge, Marschall S.
8  ? TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
9  ? NUMBER OF SEQUENCES: 12
10 ? CORRESPONDENCE ADDRESSES:
11 ? ADDRESSEE: Kilpatrick & Cody
12 ? STREET: 100 Peachtree Street
13 ? CITY: Atlanta
14 ? STATE: Georgia
15 ? COUNTRY: US
16 ? ZIP: 30303
17 ? COMPUTER READABLE FORM:
18 ? MEDIUM TYPE: Floppy disk
19 ? COMPUTER: IBM PC compatible
20 ? OPERATING SYSTEM: PC-DOS/MS-DOS
21 ? SOFTWARE: PatentIn Release #1.0, Version #1.25
22 ? CURRENT APPLICATION DATA:
23 ? APPLICATION NUMBER: US/08/212,133A
24 ? FILING DATE: March 11, 1994
25 ? CLASSIFICATION: 435
26 ? PRIOR APPLICATION DATA:
27 ? APPLICATION NUMBER: US 07/864,004
28 ? FILING DATE: 07-APR-1992
29 ? ATTORNEY/AGENT INFORMATION:
30 ? NAME: Pabst, Patricia L.
31 ? REGISTRATION NUMBER: 31,284
32 ? REFERENCE/DOCKET NUMBER: EMO/76677
33 ? TELECOMMUNICATION INFORMATION:
34 ? TELEPHONE: 404-572-6508
35 ? TELEFAX: 404-572-6555
36 ? INFORMATION FOR SEQ. ID NO. 1:
37 ? SEQUENCE CHARACTERISTICS:
38 ? LENGTH: 9009 base pairs
39 ? TYPE: nucleic acid
40 ? STRANDEDNESS: single
41 ? TOPOLOGY: linear
42 ? MOLECULE TYPE: cDNA
43 ? HYPOTHEICAL: NO
44 ? ANTI-SENSE: NO
45 ? ORIGINAL SOURCE:
46 ? ORGANISM: Homo sapien
47 ? TISSUE TYPE: Liver
48 ? FEATURE:
49 ? NAME/KEY: misc.feature (Domain Structure)
50 ? LOCATION: 5125 . . 7053
51 ? OTHER INFORMATION: /note="Equivalent to the A3-C1-C2
52 ? OTHER INFORMATION: domain"
53 ? FEATURE:
54 ? NAME/KEY: misc.feature (Domain Structure)

```



```

1 Sequence 1, Application US/08670707A
2 Patent No. 5859204
3 GENERAL INFORMATION:
4 APPLICANT: Lollar, John S.
5 TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
6 NUMBER OF SEQUENCES: 40
7 CORRESPONDENCE ADDRESSES:
8 ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
9 STREET: 5370 Manhattan Circle Suite 201
10 CITY: Boulder
11 STATE: Colorado
12 COUNTRY: USA
13 ZIP: 80303
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/670,707A
21 FILING DATE: 26-JUN-1996
22 CLASSIFICATION: 435
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: WO PCT/US94/13200
25 FILING DATE: 15-NOV-1994
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 08/212,133
28 FILING DATE: 11-MAR-1994
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 07/864,004
31 FILING DATE: 07-APR-1992
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Greenlee, Lorraine L.
34 REGISTRATION NUMBER: 27,894
35 REFERENCE/DOCKET NUMBER: 75-95F
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: 303/499-8080
38 TELEFAX: 303/499-8089
39 INFORMATION FOR SEQ ID NO: 1:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 9009 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: double
44 TOPOLOGY: not relevant
45 MOLECULE TYPE: cDNA to mRNA
46 HYPOTHEetical: NO
47 ANTI-SENSE: NO
48 ORIGINAL SOURCE:
49 ORGANISM: Homo sapiens
50 TISSUE TYPE: Liver
51 FEATURE:
52 NAME/KEY: misc_feature
53 LOCATION: 5125..7053
54 OTHER INFORMATION: /product= "Domain Structure"
55 OTHER INFORMATION: /note= "Equivalent to the A3-C1-C2 domain"
56 FEATURE:
57 NAME/KEY: misc_feature
58 LOCATION: 1..2277
59 OTHER INFORMATION: /product= "Domain Structure"
60 OTHER INFORMATION: /note= "Equivalent to the A1-A2 domain"
61 FEATURE:
62 NAME/KEY: misc_feature
63 LOCATION: 1..2277
64 OTHER INFORMATION: /product= "Domain"
65 OTHER INFORMATION: /note= "cDNA encoding human factorVIII"
66 US-08-670-707A-1

```

Query Match	48.6%;	Score 369;	DB 2;	Length 9009;
Best Local Similarity	94.2%;	Pred. No. 3,76-112;		
Matches 404;	Conservative	0;	Mismatches 25;	Indels 0;
				Gaps 0;
QY	372	actgaactccactttttttttcccaaggatcgcgtccacatgcaaatagagctct	431	

Db	107	ATTAACTTTTGGCTTCTCCAGTTGACATTTGTAGCAATATAGTCAGCAAAATGAGACTCT	166
Qy	432	ccacctgcttcttcctgtgaccttttgcatcttgcttagtgcaccagaagaactacc	491
Db	167	CCAACTGGTTCTTCTGTCGACTTTGGCATTTGCTTATGCGACACAGAAATATCACTAC	226
Qy	492	tgggtgcagtggaactcattcatgggacctatgcaaaagtatctccgttgagctgcctg	551
Db	227	TGGGCGAGTGTGAACCTGCAAGGACATATATGCAAACTGATCTGGGAGACTCCCTGTGG	286
Qy	552	acgcaaaattccctccaaagttgcacaaactcttccattccaacaccctcagtcgttaca	611
Db	287	ACGCAAGTTCCTCCCTAGATGAGTGCAGCAAAATCTTTTCCATTCAACACTCACTAGTGTATACA	346
Qy	612	aaaagacctgttctgtgaatctacaggatcaaccttttcaacatcgtctaaagccaaagccac	671
Db	347	AAAAGACTCTGTTGTATGAATTAACAGGTTCACCTTTCAACATCGCTAAACCAAGCCAC	406
Qy	672	ccctgagatggatcgcctagtcacctaccacccagctcgaagtttatagataaagtgatcat	731
Db	407	CGTGATGGTCTGCTAGTCTCTACATCCATCCAGGCTGAGGTTATGATACAGTGGTCATTAA	466
Qy	732	cacttaagaacaatggtctcccatcctctgaagttctcatgtcgttggtggtatctccactgga	791
Db	467	CACTTAAACAATGGCTTCCATCCTGTCACTCTTTCANAGCTGTGGTATCTCTACTGGA	526
Qy	792	aagctctcg 800	
Db	527	AAGCTTCTG 535	

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30309-4530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03275  
FILING DATE: 19930407  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864004  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU 106PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6508  
TELEFAX: 404-815-6555

```

? SOURCE: CHARACTERISTICS:
? LENGTH: 9009 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: single

```

```

? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE: N-terminal
? ORIGINAL SOURCE:
? ORGANISM: Homo sapien
? TISSUE TYPE: Liver

```

```
? FEATURE:  
? NAME/KEY : misc_feature  
? LOCATION : 5001..7053  
? OTHER INFORMATION: /note= "Domain Structure"  
? OTHER INFORMATION: Equivalent to the A3-C1-C2 domain"  
? FEATURE:  
? NAME/KEY : misc_feature  
? LOCATION : 1..2277
```

```

; OTHER INFORMATION: /note= "Domain Structure:
; OTHER INFORMATION: Equivalent to the A1-A2 domain"
PCT-US93-03275-3

Query Match      48.6%  Score 389;  DB 5;  Length 9009;
Best Local Similarity 94.2%  Prod No.3 7e-1123;
Matches 400;  Conservative 35;  Indels 0;  Gaps 0;

```

Accession	Organism	Gene	Size (bp)	Accession	Organism	Gene	Size (bp)
QY	<i>Acetivibrio</i>	372	372	QY	<i>Acetivibrio</i>	372	372
Dd	<i>Acetivibrio</i>	107	107	Dd	<i>Acetivibrio</i>	107	107
QY	<i>Acetivibrio</i>	432	432	QY	<i>Acetivibrio</i>	432	432
157	<i>Acetivibrio</i>	157	157	157	<i>Acetivibrio</i>	157	157

[illegible]

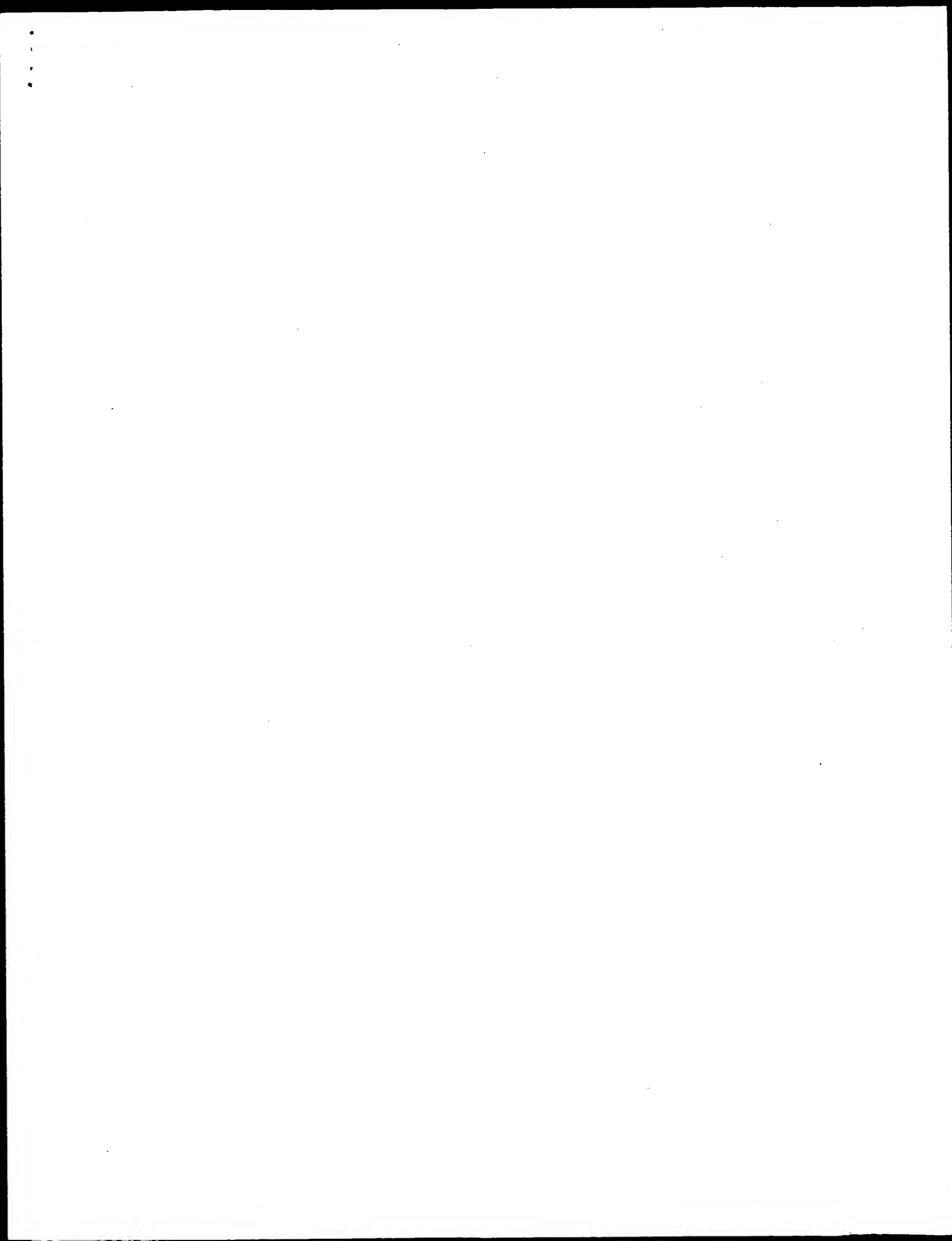


```

QY 612 aaagactctgttctgtagaatcagcagatcacccttcaacatcgctaagccaagccac 671
    |||||
Db 347 AAAGACTCTGTGTGAGAAATTCACGGTTCACCTTTCAACATCGCTAAGCCAGGCCAC 406
    |||||
QY 672 cctggatgggtctcgtactagtcctacatccaggctgagglttatgatacagtggtcatla 731
    |||||
Db 407 CCTGGATGGGTCTGCTAGGTCCTACCATCCAGGCTGAGGTTATGATACAGTGTGATTA 466
    |||||
QY 732 caactlaagaacatggtctccatccatccctgctcagtcctatgctgtgtgtatcctactga 791
    |||||
Db 467 CACTTAAGAACAATGCTTCCCATCTCTGTCAGTTCATGCTGTGTGTATCCTACTGGA 526
    |||||
QY 792 aagcttctg 800
    |||||
Db 527 AAGCTTCTG 535
    |||||

```

Search completed: June 13, 2002, 22:41:28  
 Job time: 18136 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 12:08:36 ; Search time 12.98 Seconds  
(without alignments)  
26.345 Million cell updates/sec

Title: US-09-740-211-15

Perfect score: 75

Sequence: 1 SFSQNPVLRHQR 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	100.0	14	2	US-08-634-001C-1
2	75	100.0	14	4	US-09-470-618-15
3	75	100.0	14	4	US-09-209-916-1
4	65	86.7	1661	2	US-08-882-083-2
5	65	86.7	1661	2	US-08-558-107-2
6	65	86.7	1661	4	US-09-243-539-2
7	65	86.7	2332	1	US-07-864-004B-4
8	65	86.7	2332	1	US-08-251-937A-4
9	65	86.7	2332	1	US-08-212-133A-2
10	65	86.7	2332	1	US-08-276-594A-2
11	65	86.7	2332	1	US-08-474-503-2
12	65	86.7	2332	2	US-08-670-707A-2
13	65	86.7	2332	4	US-09-037-601-2
14	65	86.7	2332	4	US-09-324-867-3
15	65	86.7	2332	5	PCT-US93-03275-4
16	65	86.7	2332	5	PCT-US94-13200-2
17	65	86.7	2332	1	US-08-121-202-2
18	65	86.7	2332	1	US-08-366-851A-2
19	65	86.7	2332	6	5171844-2
20	65	86.7	2332	6	5422260-1
21	58	77.3	1471	1	US-08-683-839B-3
22	58	74.7	2304	4	US-09-324-867-4
23	56	74.7	2319	1	US-08-212-133A-8
24	56	74.7	2319	1	US-08-474-503-6
25	56	74.7	2319	2	US-08-670-707A-6
26	56	74.7	2319	4	US-09-037-601-6
27	56	74.7	2319	5	PCT-US94-13200-6

28	55	73.3	34	3	US-08-441-935-30	Sequence 30, Appl
29	55	73.3	34	4	US-08-441-943-30	Sequence 30, Appl
30	54	72.0	2343	4	US-09-324-867-2	Sequence 2, Appl
31	50	66.7	9	4	US-08-441-943-34	Sequence 34, Appl
32	47	62.7	868	1	US-07-864-004B-6	Sequence 6, Appl
33	47	62.7	868	1	US-08-251-937A-6	Sequence 6, Appl
34	47	62.7	868	1	US-08-212-133A-3	Sequence 6, Appl
35	47	62.7	1090	5	PCT-US93-03275-6	Sequence 6, Appl
36	47	62.7	2115	4	US-09-324-867-5	Sequence 37, Appl
37	47	62.7	2133	2	US-08-670-707A-37	Sequence 37, Appl
38	47	62.7	2133	4	US-09-037-601-37	Sequence 37, Appl
39	38	50.7	364	1	US-08-318-831-6	Sequence 6, Appl
40	38	50.7	1333	3	US-09-356-952-2	Sequence 2, Appl
41	37	49.3	158	4	US-09-374-135-2	Sequence 2, Appl
42	36	48.0	274	1	US-08-248-466B-10	Sequence 12, Appl
43	36	48.0	351	1	US-08-248-466B-12	Sequence 12, Appl
44	36	48.0	392	4	US-09-416-050A-2	Sequence 2, Appl
45	36	48.0	392	4	US-09-664-800-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-634-001C-1  
; Sequence 1, Application US/08634001C  
; Patent No. 5952198  
; GENERAL INFORMATION:  
; APPLICANT: Chan, Sham-Yuen  
; TITLE OF INVENTION: Production of Recombinant Factor VIII  
; TITLE OF INVENTION: In the Presence of Liposome-Like  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSES: Bayer Corporation  
; STREET: 800 Dwight Way  
; STREET: P. O. Box 1986  
; CITY: Berkeley  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94701-1986  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB Storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Wordperfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/634,001C  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/434,900  
; FILING DATE: May 4, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gblin, James A.  
; REGISTRATION NUMBER: 25772  
; REFERENCE/DOCKET NUMBER: MSB-7226CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510)705-7910  
; TELEFAX: (510)705-7904  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14  
; TYPE: amino acid  
; STRANDEDNESS: single strand  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: peptide  
; US-08-634-001C-1  
Query Match 100.0%; Score 75; DB 2; Length 14;  
Best Local Similarity 100.0%; Pred. No. 4.2e-07;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SFSQNPVLRHOR 14  
|||||

Db 1 SFSQNPVLRHOR 14

RESULT 2  
US-09-470-618-15  
; Sequence 15, Application US/09470618  
; Patent No. 6200560  
; GENERAL INFORMATION:  
; APPLICANT: Couto, Linda B.  
; APPLICANT: Colosi, Peter C.  
; TITLE OF INVENTION: Adeno-Associated Vectors for Expression of Factor VIII  
; TITLE OF INVENTION: by Target Cells  
; FILE REFERENCE: Avigen-04082  
; CURRENT APPLICATION NUMBER: US/09/470,618  
; EARLIER FILING DATE: 1999-12-22  
; EARLIER APPLICATION NUMBER: 09/364,862  
; EARLIER FILING DATE: 1999-07-30  
; EARLIER APPLICATION NUMBER: 60/125,974  
; EARLIER FILING DATE: 1999-03-24  
; EARLIER APPLICATION NUMBER: 60/104,994  
; EARLIER FILING DATE: 1998-10-20  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-470-618-15

Query Match 100.0%; Score 75; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 4.2e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SFSQNPVLRHOR 14  
|||||

Db 1 SFSQNPVLRHOR 14

RESULT 3  
US-09-209-916-1  
; Sequence 1, Application US/09209916  
; Patent No. 6358703  
; GENERAL INFORMATION:  
; APPLICANT: Cho, Myung-Sam  
; APPLICANT: Chan, Sham Yuen  
; APPLICANT: Kelsey, William  
; APPLICANT: Yee, Helena  
; TITLE OF INVENTION: Expression System for Factor VIII  
; FILE REFERENCE: MSB-7255  
; CURRENT APPLICATION NUMBER: US/09/209,916  
; CURRENT FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1438  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Derived from  
; OTHER INFORMATION: human factor VIII sequence  
US-09-209-916-1

Query Match 100.0%; Score 75; DB 4; Length 1438;  
Best Local Similarity 100.0%; Pred. No. 6.1e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SFSQNPVLRHOR 14  
|||||

Db 741 SFSQNPVLRHOR 754

RESULT 4  
US-08-882-083-2  
; Sequence 2, Application US/08882083  
; Patent No. 5869292  
; GENERAL INFORMATION:  
; APPLICANT: VOORBERG, Johannes J.  
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/882,083  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/558,107  
; FILING DATE: 13-NOV-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ISACSON, John P.  
; REGISTRATION NUMBER: 33,715  
; REFERENCE/DOCKET NUMBER: 30472/212  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1661 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-882-083-2

Query Match 86.7%; Score 65; DB 2; Length 1661;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SQNPVLRHOR 14  
|||||

Db 966 SQNPVLRHOR 977

RESULT 5  
US-08-558-107-2  
; Sequence 2, Application US/08558107  
; Patent No. 5910481  
; GENERAL INFORMATION:  
; APPLICANT: VOORBERG, Johannes J.  
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA

ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/558,107  
FILING DATE: 13-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ISACSON, John P.  
REGISTRATION NUMBER: 33,715  
REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1661 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-558-107-2

Query Match 86.7%; Score 65; DB 2; Length 1661;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLRKHRR 14  
DB 966 SONPPVLRKHRR 977

RESULT 6  
US-09-243-539-2  
Sequence 2, Application US/09243539  
Patent No. 6130203  
GENERAL INFORMATION:  
APPLICANT: VOORBERG, Johannes J.  
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/243,539  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/558,107  
FILING DATE: 13-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ISACSON, John P.  
REGISTRATION NUMBER: 33,715  
REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1661 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-243-539-2

Query Match 86.7%; Score 65; DB 4; Length 1661;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLRKHRR 14  
DB 966 SONPPVLRKHRR 977

RESULT 7  
US-07-864-004B-4  
Sequence 4, Application US/07864004B  
Patent No. 5364771  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.  
APPLICANT: Runge, Marschall S.  
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/864,004B  
FILING DATE: 07 APRIL 1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Padst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU106  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6508  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
TISSUE TYPE: Liver CDNA sequence  
US-07-864-004B-4

Query Match 86.7%; Score 65; DB 1; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLRKHRR 14  
DB 1637 SONPPVLRKHRR 1648

RESULT 8  
US-08-251-937A-4  
; Sequence 4, Application US/08251937A  
; Patent No. 5583209  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; APPLICANT: Runge, Marshall S.  
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 1100 Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: US  
; ZIP: 30309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/251,937A  
; FILING DATE: 31-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/864,004  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Patel, John S.  
; REGISTRATION NUMBER: 29,476  
; REFERENCE/DOCKET NUMBER: EMU106DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-815-6367  
; TELEFAX: 404-815-6555  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2332 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; TISSUE TYPE: Liver CDNA sequence  
; US-08-251-937A-4

Query Match 86.7%; Score 65; DB 1; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPVLKRHR 14  
|||||  
Db 1637 SONPVLKRHR 1648

RESULT 9  
US-08-212-133A-2  
; Sequence 2, Application US/08212133A  
; Patent No. 5663060  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; APPLICANT: Runge, Marshall S.  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 100 Peachtree Street  
; CITY: Atlanta

STATE: Georgia  
COUNTRY: US  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/212,133A  
FILING DATE: March 11, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,004  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU/76677  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-572-6508  
TELEFAX: 404-572-6555  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
TISSUE TYPE: Liver CDNA sequence  
US-08-212-133A-2

Query Match 86.7%; Score 65; DB 1; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPVLKRHR 14  
|||||  
Db 1637 SONPVLKRHR 1648

RESULT 10  
US-08-276-594A-2  
; Sequence 2, Application US/08276594A  
; Patent No. 5693499  
; GENERAL INFORMATION:  
; APPLICANT: YONEMURA, Hiroshi  
; APPLICANT: TAJIMA, Yoshitaka  
; APPLICANT: SUGAWARA, Keishin  
; APPLICANT: MASUDA, Kenichi  
; TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION  
; TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,594A

FILING DATE: 18-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/950,191  
FILING DATE: 24-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 243262/1991  
FILING DATE: 24-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/195/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-276-594A-2

Query Match 86.7%; Score 65; DB 1; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 SONPVYLRHOR 14  
|||||  
DB 1637 SONPVYLRHOR 1648

RESULT 11  
US-08-474-503-2  
Sequence 2, Application US/08474503  
Patent No. 5744446  
GENERAL INFORMATION:  
APPLICANT: Emory University  
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: US  
ZIP: 30309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,503  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pratt, John S.  
REGISTRATION NUMBER: 29,476  
REFERENCE/DOCKET NUMBER: EMU106CIP(3)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6500  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES

ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapien  
TISSUE TYPE: Liver cDNA sequence  
US-08-474-503-2

Query Match 86.7%; Score 65; DB 1; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 SONPVYLRHOR 14  
|||||  
DB 1637 SONPVYLRHOR 1648

RESULT 12  
US-08-670-707A-2  
Sequence 2, Application US/08670707A  
Patent No. 5859204  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.  
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/670,707A  
FILING DATE: 26-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US94/13200  
FILING DATE: 15-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,133  
FILING DATE: 11-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,004  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenlee, Lorraine L.  
REGISTRATION NUMBER: 27,894  
REFERENCE/DOCKET NUMBER: 75-95F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Liver  
US-08-670-707A-2



Query Match 86.7%; Score 65; DB 2; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLKRHR 14  
|||||  
DB 1637 SONPPVLKRHR 1648

## RESULT 13

US-09-037-601-2  
; Sequence 2, Application US/09037601  
; Patent No. 6180371  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/037,601  
; FILING DATE: 26-JUN-1996  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US94/13200  
; FILING DATE: 15-NOV-1994  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/212,133  
; FILING DATE: 11-MAR-1994  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/864,004  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Felder, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 75-95F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/499-8080  
; TELEFAX: 303/499-8089  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2332 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: YES  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Liver  
; US-09-037-601-2

Query Match 86.7%; Score 65; DB 4; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLKRHR 14  
|||||  
DB 1637 SONPPVLKRHR 1648

RESULT 14  
US-09-324-867-3

; Sequence 3, Application US/09324867A  
; Patent No. 6251632  
; GENERAL INFORMATION:  
; APPLICANT: Lilliecrap, David  
; APPLICANT: Cameron, Cherie  
; APPLICANT: No. 6251632ley, Colleen  
; APPLICANT: Horrocks, L. Suzanne Hoyle  
; APPLICANT: Hough, Christine  
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use  
; FILE REFERENCE: 1669.0010002/JAG/BJD  
; CURRENT APPLICATION NUMBER: US/09/324,867A  
; CURRENT FILING DATE: 1999-06-03  
; EARLIER APPLICATION NUMBER: 09/035,141  
; EARLIER FILING DATE: 1998-03-05  
; EARLIER APPLICATION NUMBER: 60/039,953  
; EARLIER FILING DATE: 1997-03-06  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Patentlin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-324-867-3

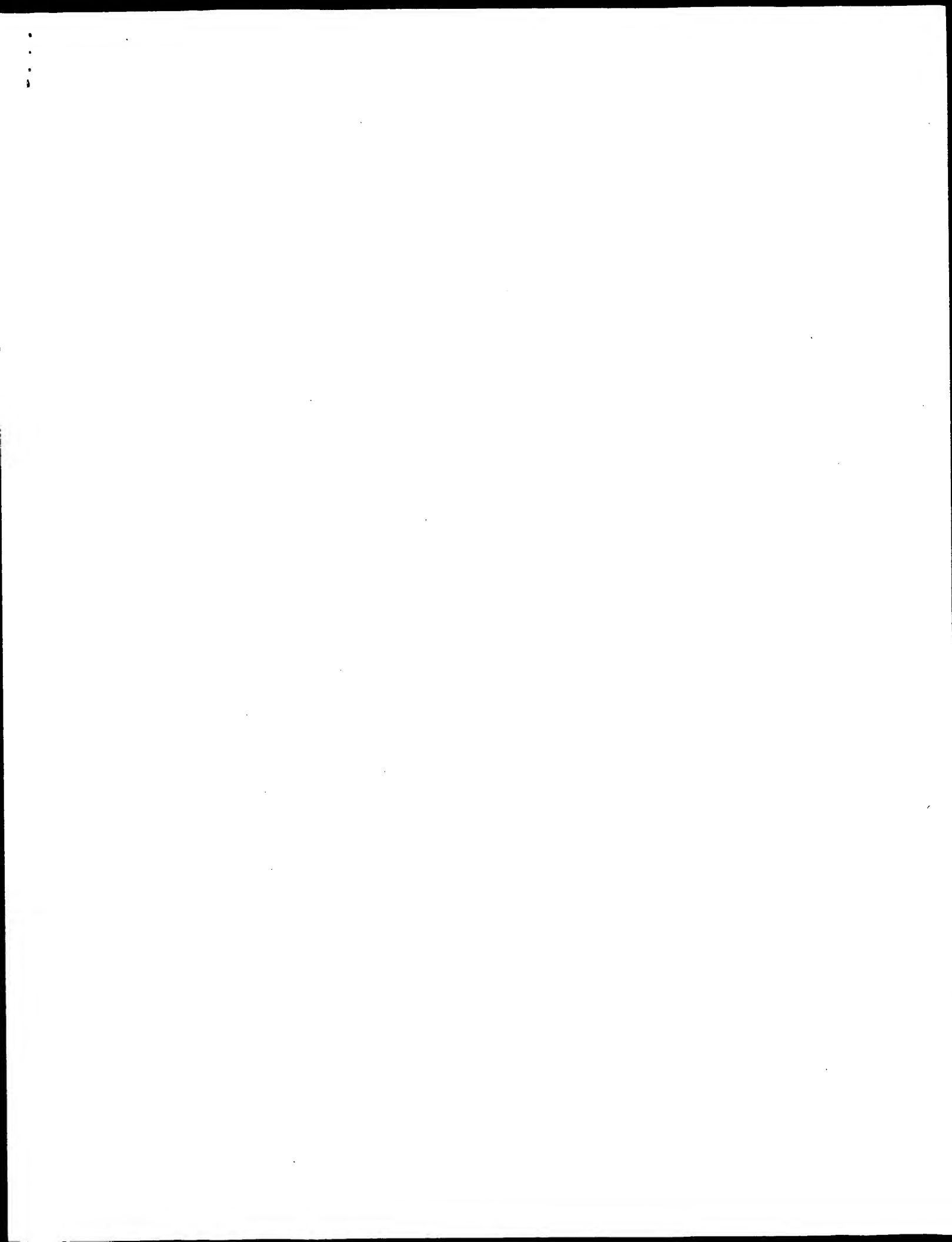
Query Match 86.7%; Score 65; DB 4; Length 2332;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 12: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLKRHR 14  
|||||  
DB 1637 SONPPVLKRHR 1648

## RESULT 15

PCT-US93-03275-4  
; Sequence 4, Application PC/TUS9303275  
; GENERAL INFORMATION:  
; APPLICANT: Lollar, John S.  
; APPLICANT: Runge, Marshall S.  
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 1100 Peachtree Street, Suite 2800  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: US  
; ZIP: 30309-4530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentlin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/03275  
; FILING DATE: 19930407  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/864004  
; FILING DATE: 07-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: EMU 106PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404-815-6508  
; TELEFAX: 404-815-6555  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 12:09:16 ; Search time 29.99 Seconds  
(without alignments)  
51.852 Million cell updates/sec

Title: US-09-740-211-15

Perfect score: 75  
Sequence: 1 SFSQNPVLRHQR 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A-Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	100.0	14	AAW04651	Peptide linker for
2	75	100.0	14	AAE04332	Human factor VIII
3	75	100.0	1438	AAE01262	B-domain deleted f
4	75	100.0	1440	AAI12971	Factor VIII:SO.
5	75	100.0	1457	AAW46246	Human factor VIII
6	75	100.0	1457	AAW44372	Human factor VIII
7	75	100.0	1457	AAW21675	Beta-domain delete
8	65	86.7	36	AAW56776	Human anti-hemoph
9	65	86.7	86	AAW56775	Human anti-hemoph
10	65	86.7	211	AAW50091	truncated derivati
11	65	86.7	211	AAW07203	Human Factor VIIIc

12	65	86.7	1283	21	AAW07205	Human Factor VIIIc
13	65	86.7	1284	6	AAW50108	Factor-VIII deriv
14	65	86.7	1516	9	AAW80265	Modified factor VI
15	65	86.7	1661	18	AAW18670	Factor VIII-d8695-
16	65	86.7	2098	17	AAW86863	Factor-VIII. Homo
17	65	86.7	2332	8	AAW71727	Factor VIII:c varia
18	65	86.7	2332	8	AAW71728	Factor VIII:c varia
19	65	86.7	2332	8	AAW71729	Factor VIII:c varia
20	65	86.7	2332	14	AAW43257	Human Factor VIII.
21	65	86.7	2332	18	AAW33222	Procoagulant-activ
22	65	86.7	2332	18	AAW33223	Procoagulant-activ
23	65	86.7	2332	18	AAW33224	Procoagulant-activ
24	65	86.7	2332	18	AAW33225	Procoagulant-activ
25	65	86.7	2332	18	AAW33226	Procoagulant-activ
26	65	86.7	2332	19	AAW53483	Human factor VIII.
27	65	86.7	2332	19	AAW44132	Homo sapiens modif
28	65	86.7	2332	20	AAW31594	Human factor VIII
29	65	86.7	2332	21	AAW57847	Human factor VIII
30	65	86.7	2332	22	AAE10826	Human mature wild-
31	65	86.7	2332	22	AAE11200	N-terminal truncat
32	65	86.7	2332	22	AAW71902	Human factor VIII.
33	65	86.7	2332	22	AAW50465	Human factor VIII.
34	65	86.7	2342	18	AAW11422	Active Factor VIII
35	65	86.7	2342	18	AAW11349	Active Factor VIII
36	65	86.7	2342	18	AAW11393	Active Factor VIII
37	65	86.7	2343	18	AAW11335	Active Factor VIII
38	65	86.7	2344	18	AAW11432	Active Factor VIII
39	65	86.7	2344	18	AAW11410	Active Factor VIII
40	65	86.7	2344	18	AAW11384	Active Factor VIII
41	65	86.7	2344	18	AAW11361	Active Factor VIII
42	65	86.7	2344	18	AAW11367	Active Factor VIII
43	65	86.7	2344	18	AAW11370	Active Factor VIII
44	65	86.7	2344	18	AAW11331	Active Factor VIII
45	65	86.7	2345	18	AAW11466	Active Factor VIII

#### ALIGNMENTS

RESULT 1	
ID	AAW04651 standard; peptide; 14 AA.
AC	AAW04651;
DT	06-AUG-1997 (first entry)
DE	Peptide linker for truncated factor VIII 90 and 80 kd fragments.
KW	linker; cell culture medium; liposome; lipid; phosphatidylcholine;
KW	phosphatidylserine; increase; production; recombinant factor VIII;
KW	truncated; haemophilia; treatment; plasma protein.
OS	Synthetic.
XX	
PN	EP745672-A2.
XX	
PD	04-DEC-1996.
XX	
PF	25-APR-1996; 96EP-0106482.
XX	
PR	17-APR-1996; 96US-0634001.
XX	
PR	04-MAY-1995; 95US-0434900.
XX	
PA	(FARB ) BAYER CORP.
PA	(CHAN/) CHAN S.
PA	(MILE ) MILES INC.
XX	
PI	Chan S;
XX	
DR	WPI; 1997-013695/02.
XX	
PT	Culture medium contg. liposome like substance comprising at least 2

PT different lipid(s) - useful for increasing prodn. of recombinant  
PT Factor VIII in mammalian cell cultures

PS Claim 15; Page 7; 7pp; English.

XX A novel cell culture medium contg. a liposome like substance, comprises  
CC at least 2 different lipids in a molar ratio sufficient to assure a  
CC 4-fold increase in factor VIII expression in a mammalian cell culture  
CC system. The culture medium is useful to increase production of  
CC recombinant factor VIII 4-fold and truncated factor VIII 3-fold.  
CC Factor VIII is a plasma protein required for clotting of the blood,  
CC useful in treatment of haemophilia. Liposome like substances contg.  
CC lipids such as phosphatidylcholine (PC), phosphatidylethanolamine (PE)  
CC or phosphatidylserine (PS) alone have no effect on recombinant factor  
CC VIII expression in BHK-21 and 293S cells. The present sequence is a  
CC linker joining truncated recombinant factor VIII 90 and 80 kd  
CC fragments.

SQ Sequence 14 AA;

Query Match 100.0%; Score 75; DB 18; Length 14;  
Best Local Similarity 100.0%; Pred. No. 6.9e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 1 SFSQNPPVLRKHOR 14  
Db 1 sfsgnpvllkrhgr 14  
|||||  
|||

RESULT 2  
AAE04332 ID AAE04332 standard; peptide; 14 AA.  
XX AC AAE04332:  
XX DF 04-SEP-2001 (first entry)  
DE Human factor VIII B-domain partial sequence.  
KW Recombinant adeno-associated viral vector; rAAV; factor VIII; FVIII;  
KM blood clotting disorder; gene therapy; haemophilia A; human; B-domain.  
XX Homo sapiens.  
OS WO200145510-A1.  
XX PN 28-JUN-2001.  
XX PD 21-DEC-2000; 2000MO-US34925.  
XX PF 22-DEC-1999; 99US-0470618.  
XX PR (AVIG-) AVIGEN INC.  
PA Couto LB, Colosi PC, Qian X;  
PI WPI; 2001-417955/44.  
DR Treating blood clotting disorder, especially hemophilia in mammals, by  
PT administering recombinant adeno-associated vectors which express blood  
PT coagulation factor VIII -  
PS Claim 6; Fig 2; 90pp; English.

XX The present invention relates to a method for treating a subject  
CC suffering from a blood clotting disorder. The method comprises  
CC administering a recombinant adeno-associated virion (rAAV) comprising  
CC a nucleotide sequence encoding the light chain of factor VIII and a  
CC second recombinant adeno-associated virion comprising a nucleotide  
CC sequence encoding the heavy chain of factor VIII. The rAAV vector is  
CC useful in gene therapy for treating haemophilia A in mammals, in  
CC particular humans. The rAAV vector provides high level and long term

CC	expression of biologically active clotting factor VIII in vivo.
CC	The present sequence represents partial human factor VIII B-domain.
CC	This sequence is encoded by rAAV vectors of the invention.
XX	
SO	Sequence 14 AA:
	Query Match 100.0%; Score 75; DB 22; Length 14;
	Best Local Similarity 100.0%; Pred. NO. 6.9e-07;
	Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 SFSQNPVLYKRHOR 14
Db	1 sfsgnpylkrhqr 14
RESULT 3	
AA01262	
ID	AA01262 standard; protein: 1438 AA.
XX	
AC	AA01262;
XX	
DT	25-SEP-2000 (first entry)
XX	
DE	B-domain deleted factor VIII sequence.
XX	
KW	Factor VIII; procoagulant; adenovirus; adeno-associated strain;
KW	gene therapy; human Burkitt's lymphoma; HKB; therapy;
KW	therapeutic protein; vector; Epstein-Barr virus; human.
XX	
OS	Homo sapiens.
XX	
PN	WO20034505-A1.
XX	
PD	15-JUN-2000.
XX	
PF	08-DEC-1999; 99WO-US29169.
XX	
PR	10-DEC-1998; 98US-0209916.
XX	
PA	(FARB ) BAYER CORP.
XX	
PI	Cho M, Chan SY, Kelsey W, Yee H;
XX	
DR	WPI; 2000-431311/37.
XX	
PT	Producing cells expressing a protein having factor VIII procoagulant
PT	activity especially, human factor VIII in an industrial scale, involves
PT	expressing a vector comprising a sequence coding for factor VIII in
PT	human cells
XX	
PS	Claim 7; Fig 1; 27pp; English.
XX	
CC	Producing cells expressing a protein having factor VIII procoagulant
CC	activity, comprises contacting the cells with a vector comprising a
CC	selectable marker and a sequence coding for the protein having factor
CC	VIII procoagulant activity operably linked to a promoter. The cells
CC	are then selected and individual clones expressing high levels of the
CC	protein are isolated from the selected cells. The cells produced by
CC	the method are not only useful for producing protein having factor
CC	VIII procoagulant activity but also for producing adenovirus and
CC	adeno-associated virus strains for gene therapy. The advantage of
CC	having cells producing protein with factor VIII procoagulant activity
CC	is that factor VIII protein can be produced on an industrial scale
CC	in the range of 2-4 pg/cell/day. Human Burkitt's lymphoma (HKB)
CC	cells provide a protein-free production system to produce not only
CC	B-domain deleted factor VIII but also other therapeutic proteins. The
CC	vector used in the method preferably comprises B-domain deleted
CC	factor VIII (BDD-FVIII), a transcriptional unit for BDD-FVIII and a
CC	selectable marker, dihydrofolate reductase (dhfr). In addition, a
CC	terminal repeat sequence from Epstein-Barr virus is inserted into the
CC	vector to increase integration efficiency.
XX	



antitrypsin deficiency, Gaucher's syndrome, anaemia, infections such as HIV infection, high blood pressure, Alzheimer's disease, autoimmune or inflammatory disease or graft versus host disease. RV's are capable of surviving inactivation in human serum thereby allowing efficient gene transfer over prolonged periods of time.

Sequence 1457 AA;

Query Match 100.0%; Score 75; DB 19; Length 1457;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SFSQNPVLRKHQR 14  
DB 760 sfsqnpvlrkhr 773

RESULT 6

AAW4372 standard; Protein: 1457 AA.

AAW4372;

20-JUL-1998 (first entry)

Human Factor VIII SQN deletion mutant.

Factor VIII; blood clotting; haemophilia A; gene therapy;

retrovirus; vector; human.

Homo sapiens.

Synthetic.

MO9800542-A2.

08-JAN-1998.

02-JUL-1997; 97WO-US11785.

04-JUN-1997; 97US-0869309.

03-JUL-1996; 96US-0645601.

13-AUG-1996; 96US-0636381.

(CHIR ) CHIRON CORP.

Allen JR, Barber JR, Boder M, Chang SM, Chong K;

De la Vega D, Depolo NT, Greengard J, Hsu DC, Ibanez CE;

Jolly DJ, Mittelstaedt DM, Prussak CE, Respass JG;

WPI; 1998-086967/08.

N-PSDB; AAV15338.

New replication defective recombinant retroviruses - which express B

domain-deleted human factor VIII or human factor IX for the

treatment of haemophilia

Claim 5; Page 175-180; 236pp; English.

This polypeptide comprises the B domain deletion mutant SQN of human factor VIII. The SQN mutant is created by fusing Ser-743 to Gln-1638 of native factor VIII (see AAW4373) to form a Ser-Gln-Asn (SQN) link between the A2 and A3 Factor VIII domains. A DNA sequence encoding the SQN deletion mutant is provided in AAV15338. When compared to plasmid-derived factor VIII, the SQN deletion does not influence the in vivo pharmacokinetics, but the reduced size of the molecule appears to decrease proteolytic degradation. The invention relates to preparations of replication defective recombinant retrovirus (RV) expressing a B domain-deleted human factor VIII protein, where the recombinant RV is capable of infecting human cells, is resistant to degradation by human complement and is capable of inducing long-term (at least 30 days and up to 6 months or longer post-injection) systemic expression of

Factor VIII when administered to a haemophilia A patient.

Sequence 1457 AA;

Query Match 100.0%; Score 75; DB 19; Length 1457;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SFSQNPVLRKHQR 14  
DB 760 sfsqnpvlrkhr 773

RESULT 7

AAV21675 standard; Protein: 1457 AA.

AAV21675;

18-AUG-1999 (first entry)

Beta-domain deleted factor VIII protein.

Factor VIII protein; gene modification; gene therapy; clinical disorder;

splicing pattern; RNA processing; gene regulation; beta-domain; human.

Homo sapiens.

WO9929848-A1.

17-JUN-1999.

25-NOV-1998; 98WO-US25354.

16-JAN-1998; 98US-0071596.

05-DEC-1997; 97US-0067614.

(IMMO-) IMMUNE RESPONSE CORP.

Bidlinauer S, Gonzales JEN, Ill CR, Yang CO;

WPI; 1999-385602/32.

N-PSDB; AAX82258, AAX82259, AAX82260.

Genes and vectors exhibiting increased expression and novel splicing

patterns, useful for expression of, e.g. beta-domain deleted factor

VIII

Disclosure; Page 72-78; 123pp; English.

The invention describes novel genes and vectors exhibiting increased expression and novel splicing patterns. It provides a gene encoding a factor VIII protein, that comprises one or more consensus or near consensus splice sites which have been corrected to increase expression. The method, DNA sequences and expression vectors can be used to increase the expression of a gene, especially a factor VIII gene. Genes containing modified 5' and/or 3' untranslated regions have optimized expression levels and tissue-specific expression. The methods are used for identification and correction of consensus splice sites, addition of introns, optimization of 5' and 3' untranslated regions and increase in cytoplasmic RNA accumulation. Hence the DNAs are useful in gene therapy to treat a clinical disorder. To study RNA processing and/or gene regulation. The present sequence represents a beta-domain deleted factor VIII protein.

Sequence 1457 AA;

Query Match 100.0%; Score 75; DB 20; Length 1457;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 SFSQNPVLKRRHOR 14  
 DB 760 SFSQNPVLKRRHOR 773

## RESULT 8

AAR56776  
 ID AAR56776 standard; peptide: 36 AA.

XX AAR56776;

XX 03-APR-1995 (first entry)

DE Human anti-haemophilic factor VIII B domain C-terminal fragment.

XX Anthaemophilic factor VIII; B domain; monoclonal antibody; Mab;

KW fusion protein; impurity; immuno-adsorption.

XX Homo sapiens.

OS JP06205696-A.

XX 26-JUL-1994.

XX 11-JAN-1993; 93JP-0002537.

XX 11-JAN-1993; 93JP-0002537.

XX (TEIJU) TEIJUN LTD.

PA (KAGA) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.

XX WPI: 1994-275529/34.

DR Monoclonal antibody binding to unnatural human anti-haemophilic

PT factor VIII B domain - is used in immuno-adsorption process to

XX bind impurities

PS Claim 3; Page 10; 12pp; Japanese.

CC A monoclonal antibody (Mab) directed against this peptide may be

CC used in an immuno-adsorption process to remove unnatural type human

CC antihaemophilic factor VIII L chain having this peptide fused to its

CC N-terminus (an impurity on preparation of the human antihaemophilic

CC factor VIII protein complex). The Mab binds to unnatural type human

CC antihaemophilic factor VIII L chain fused with the peptide but does

CC not bond substantially to natural type human antihaemophilic factor

CC VIII L chain. This peptide corresponds to amino acids 1614-1649 of

CC human antihaemophilic factor VIII-B domain.

QY 3 SQNPVLKRRHOR 14

DB 24 SQNPVLKRRHOR 35

## RESULT 9

AAR56775  
 ID AAR56775 standard; peptide: 86 AA.

XX AAR56775;

XX 03-APR-1995 (first entry)

DE Human anti-haemophilic factor VIII B domain C-terminal fragment.

XX Anthaemophilic factor VIII; B domain; monoclonal antibody; Mab;

KW fusion protein; impurity; immuno-adsorption.

XX Homo sapiens.

OS JP06205696-A.

XX 26-JUL-1994.

XX 11-JAN-1993; 93JP-0002537.

XX 11-JAN-1993; 93JP-0002537.

XX (TEIJU) TEIJUN LTD.

PA (KAGA) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.

XX WPI: 1994-275529/34.

DR Monoclonal antibody binding to unnatural human anti-haemophilic

PT factor VIII B domain - is used in immuno-adsorption process to

XX bind impurities

PS Claim 3; Page 10; 12pp; Japanese.

CC A monoclonal antibody (Mab) directed against this peptide may be

CC used in an immuno-adsorption process to remove unnatural type human

CC antihaemophilic factor VIII L chain having this peptide fused to its

CC N-terminus (an impurity on preparation of the human antihaemophilic

CC factor VIII protein complex). The Mab binds to unnatural type human

CC antihaemophilic factor VIII L chain fused with the peptide but does

CC not bond substantially to natural type human antihaemophilic factor

CC VIII L chain. This peptide corresponds to the 86 C-terminal amino

CC acids of human antihaemophilic factor VIII-B domain (amino acids

CC 1563-1648 of the mature protein).

QY 3 SQNPVLKRRHOR 14

DB 75 SQNPVLKRRHOR 86

## RESULT 10

AAP50091  
 ID AAP50091 standard; Protein: 211 AA.

XX AAP50091;

XX 10-OCT-1991 (first entry)

DE Truncated derivative of Factor-VIIIIC.

XX Factor-VIIIIC; blood-clotting.

XX Homo sapiens.

OS EP150735-A.

XX 07-AUG-1985.

XX 11-JAN-1985; 85EP-0100223.

XX 26-OCT-1984; 84US-0664919.

XX 12-JAN-1984; 84US-0570062.

XX (CHIR-) CHIRON CORP.

PA (NORD-) NORDISK GENTOFTE.

XX Kuo G, Masierz F, Truett M, Valenzuela P, Rasmussen M;

PI Favaloro J;

XX WPI; 1985-191681/32.  
 DR N-PSDB; AAN50105.  
 XX  
 PT Prodn. of human Factor VIII C or its precursors or sub-units - by  
 PT using recombinant DNA techniques with Factor VIII C gene  
 PT expressed in host.  
 XX  
 PS Disclosure; Page 32; 58pp; English.  
 XX  
 CC The sequence encodes a truncated derivative of human Factor-VIIIc,  
 CC which is useful as an anticoagulant for the treatment of  
 CC haemophilia.  
 XX  
 SQ Sequence 211 AA;

Query Match 86.7%; Score 65; DB 6; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 0.00082;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 SONPVKRRHOR 14  
 |||||  
 Db 115 sqnpvklrhqr 126

RESULT 11  
 AAB07203  
 ID AAB07203 standard; Protein; 211 AA.  
 AC AAB07203;  
 XX  
 DT 11-OCT-2000 (first entry)  
 XX  
 DE Human Factor VIIIc 77/80kd subunit.  
 XX  
 KW Factor VIII C domain; human; blood clotting; haemophilia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1006182-A2.  
 XX  
 PD 07-JUN-2000.  
 XX  
 PF 11-JAN-1985; 2000EP-0200860.  
 XX  
 PR 12-JAN-1984; 84US-0570062.  
 PR 26-OCT-1984; 84US-0664919.  
 PR 11-JAN-1985; 85EP-0100223.  
 PR 11-JAN-1985; 91EP-0113267.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Kuo G, Rasmussen ME, Maslarsz FR, Valenzuela P, Truett M;  
 PI Favaloro J;  
 XX  
 DR WPI; 2000-367967/32.  
 DR N-PSDB; AAS58440.  
 XX  
 PT DNA sequence of portion of human Factor VIIIc for treating and  
 PT preventing the symptoms of haemophilia -  
 XX  
 PS Disclosure; Page 16; 39pp; English.  
 XX  
 CC The present sequence is the protein sequence for the 77/80kd subunit  
 CC of human Factor VIIIc. The sequence was identified by creating and  
 CC sequencing a genomic DNA library of the Factor VIIIc gene. Factor VIIIc  
 CC is a plasma protein involved in blood coagulation, and is absent or  
 CC defective in haemophilia A. The Factor VIIIc protein can, therefore, be  
 CC used to treat haemophilia, as well as in the production of monoclonal  
 CC antibodies to Factor VIIIc, and in diagnostic assays for the presence of  
 CC Factor VIIIc subunits in physiological fluids, for example blood or

CC serum.  
 XX  
 SQ Sequence 211 AA;

Query Match 86.7%; Score 65; DB 21; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 0.00082;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 SONPVKRRHOR 14  
 |||||  
 Db 115 sqnpvklrhqr 126

RESULT 12  
 AAB07205  
 ID AAB07205 standard; Protein; 1283 AA.  
 AC AAB07205;  
 XX  
 DT 11-OCT-2000 (first entry)  
 XX  
 DE Human Factor VIIIc protein sequence.  
 XX  
 KW Factor VIII C domain; human; blood clotting; haemophilia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1006182-A2.  
 XX  
 PD 07-JUN-2000.  
 XX  
 PF 11-JAN-1985; 2000EP-0200860.  
 XX  
 PR 12-JAN-1984; 84US-0570062.  
 PR 26-OCT-1984; 84US-0664919.  
 PR 11-JAN-1985; 85EP-0100223.  
 PR 11-JAN-1985; 91EP-0113267.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Kuo G, Rasmussen ME, Maslarsz FR, Valenzuela P, Truett M;  
 PI Favaloro J;  
 XX  
 DR WPI; 2000-367967/32.  
 DR N-PSDB; AAS58443.  
 XX  
 PT DNA sequence of portion of human Factor VIIIc for treating and  
 PT preventing the symptoms of haemophilia -  
 XX  
 PS Disclosure; Page 31-35; 39pp; English.  
 XX  
 CC The present sequence is the protein sequence for human Factor VIIIc,  
 CC which was obtained by sequencing a genomic DNA library. Factor VIIIc is a  
 CC plasma protein involved in blood coagulation, and is absent or defective  
 CC in haemophilia A. The Factor VIIIc protein can, therefore, be used to  
 CC treat haemophilia, as well as in the production of monoclonal antibodies  
 CC to Factor VIIIc, and in diagnostic assays for the presence of Factor  
 CC VIIIc subunits in physiological fluids, for example blood or serum.  
 XX  
 SQ Sequence 1283 AA;

Query Match 86.7%; Score 65; DB 21; Length 1283;  
 Best Local Similarity 100.0%; Pred. No. 0.0058;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 SONPVKRRHOR 14  
 |||||  
 Db 952 sqnpvklrhqr 963

```

RESULT 13
AAP50108
ID AAP50108 standard; Protein: 1284 AA.
XX
AC AAP50108;
XX
DT 10-OCT-1991 (first entry)
XX
DE Factor-VIIIc derivative.
XX
KM Factor-VIIIc; blood-clotting.
XX
OS Homo sapiens.
XX
PN EPI50735-A.
XX
PD 07-AUG-1985.
XX
PF 11-JAN-1985; 85EP-0100223.
XX
PR 26-OCT-1984; 84US-0664919.
XX
PR 12-JAN-1984; 84US-0570062.
XX
PA (CHIR-) CHIRON CORP.
XX
PA (NORD-) NORDISK GENTOFTE.
XX
PI Kuo G, Maslarz F, Truett M, Valenzuela P, Rasmussen M;
PI Favaloro J;
XX
DR WPI: 1985-191681/32.
XX
PT Prodn. of human Factor VIII C or its precursors or sub-units - by
PT using recombinant DNA techniques with Factor VIII C gene
XX expressed in host.
XX
PS Disclosure; Appendix B: 58pp; English.
XX
CC The sequence encodes a derivative of human Factor-VIIIc, which is
CC useful as an anticoagulant for the treatment of haemophilia.
XX
SQ Sequence 1284 AA;

Query Match 86.7%; Score 65; DB 6; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SQNPVVKRRQR 14
   |||||
DB 952 sqnpvklrhnqr 963

RESULT 14
AAP80265
ID AAP80265 standard; protein: 1516 AA.
XX
AC AAP80265;
XX
DT 10-OCT-1990 (first entry)
XX
DE Modified factor VIII:C sequence with the Q744-DI563 deletion.
XX
KM Modified factor VIII:C; maturation polypeptide; haemophilia;
KM blood coagulation; QD deletion.
XX
OS Homo sapiens.
XX
PN W08800831-A.
XX
PD 11-FEB-1988.
XX
PF 31-JUL-1987; 87MO-US01814.
XX

```

```

PR 01-AUG-1986; 86US-0893375.
XX
PA (BIOJ ) BIOGEN NV (PASE/).
XX
PI Pasek MP.
XX
DR WPI: 1988-049866/07.
XX
DR N-PSDB; AAN80444.
XX
PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
PT encoding maturation polypeptide, useful for high yield transformation.
XX
PS Claim 3; Page 51-52-53-54; 97pp; English.
XX
CC A major part of the sequence encoding the maturation polypeptide of
CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1363. The QD deletion
CC retains approximately 90 amino acids of the maturation polypeptide
CC (four amino acids at the N-terminal end and 86 amino acids at
CC the C-terminal end). The full length Factor VIII:C cDNA has two
CC changes with respect to the published sequence (EPO application 160457):
CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
CC (Phe to Leu).
CC The product is produced in approx. 20 times higher
CC yields than previous recombinant produced factor VIII:C and are more
CC easily purified. The peptide is used for treating haemophilia A, both
CC acute and prolonged bleeding.
CC See also AAN80446 and AAN80447.
XX
SQ Sequence 1516 AA;

```

```

Query Match 86.7%; Score 65; DB 9; Length 1516;
Best Local Similarity 100.0%; Pred. No. 0.0069;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SQNPVVKRRQR 14
   |||||
DB 821 sqnpvklrhnqr 832

RESULT 15
AAW18670
ID AAW18670 standard; Protein: 1661 AA.
XX
AC AAW18670;
XX
DT 10-AUG-1997 (first entry)
XX
DE Factor VIII-dB695-HCII.
XX
KM Factor VIII-dB695-HCII; heparin cofactor II; blood coagulation;
KM blood clotting; procoagulant; anticoagulant; antithrombotic;
KM haemophilia; gene therapy.
XX
OS Homo sapiens.
XX
PN Key
XX FH Location/Qualifiers
XX FT 731..760
XX FT //label= "HCII
XX FT //note= "heparin cofactor II acidic domain"
XX
PN W09718315-A1.
XX
PD 22-MAY-1997.
XX
PF 13-NOV-1996; 96MO-EP04977.
XX
PR 13-NOV-1995; 95US-0558107.
XX
PA (IMMO ) IMMUNO AG.
XX
PI Voorberg JJ;
XX

```

DR WPI; 1997-289291/26.  
DR N-PSDB; AAT69811.

XX Hybrid Factor VIII with modified activity, comprises region from  
PT donor anticoagulant or antithrombotic protein - useful for treatment  
PT of coagulation disorders

XX Claim 11; Page 52-60; 96pp; English.

XX Factor VIII-dB695-HCII (AAW18670) is a hybrid protein in which amino  
CC acids 712-736 of Factor-dB695 (Factor VIII del1868-1562) B-domain  
CC are replaced by amino acids 51-80 from the acidic region (and  
CC potential thrombin-binding site) of human heparin cofactor II  
CC (HCII). It is the expression product of Factor VIII-dB695-HCII  
CC DNA (AAT69811) in plasmid pCIB-dB695-HCII. The hybrid protein, which  
CC can be expressed using gene therapy techniques, has increased  
CC procoagulant activity owing to the HCII acidic region, and can be  
CC used to treat blood coagulation disorders such as haemophilia A.  
XX  
SQ Sequence 1661 AA;

Query Match 86.7%; Score 65; DB 18; Length 1661;

Best Local Similarity 100.0%; Pred. No. 0.0076;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 3 SQNPPVLRHQR 14  
|||||  
Db 966 sqnppvllrhqr 977

Search completed: June 13, 2002, 12:12:51  
Job time: 215 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 12:11:42 ; Search time 14.86 Seconds  
(without alignments)

90.528 Million cell updates/sec

Title: US-09-740-211-15

Perfect score: 75

Sequence: 1 SFSQNPVLRHQR 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR.71:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	86.7	2351	1 EZHU	coagulation factor
2	56	74.7	2319	2 A47004	coagulation factor
3	47	62.7	869	2 A25945	coagulation factor
4	47	62.7	2133	2 T42763	coagulation factor
5	42	56.0	1077	2 S70120	ATP-dependent RNA
6	41	54.7	188	2 B87329	conserved hypotet
7	41	54.7	2415	1 A33733	spectrin alpha cha
8	40	53.3	423	2 T40224	protein kinase - f
9	39	52.0	506	2 G90430	conserved hypotet
10	39	52.0	697	2 JN0602	rib protein - Esch
11	39	52.0	3085	2 T00327	polyprotein - infe
12	39	52.0	3335	2 H81702	adherence factor T
13	38	50.7	131	2 B90332	hypothetical prote
14	38	50.7	199	1 S50398	ribosomal protein
15	38	50.7	199	2 S67618	ribosomal protein
16	38	50.7	269	2 T31318	hypothetical prote
17	38	50.7	303	2 T06981	low-molecular-weig
18	38	50.7	304	2 T06505	glutinin low molec
19	38	50.7	332	2 G83066	conserved hypotet
20	38	50.7	347	2 S30921	ferredoxin-nitrit
21	38	50.7	436	2 F71358	hypothetical prote
22	38	50.7	604	2 T37994	probable splicing
23	38	50.7	623	2 T15510	hypothetical prote
24	38	50.7	1144	2 T27408	hypothetical prote
25	38	50.7	1307	2 T25563	hypothetical prote
26	38	50.7	1333	2 A37488	Ras guanine nucle
27	37	49.3	251	2 A99447	hypothetical prote
28	37	49.3	251	2 S01982	glutinin low molec
29	37	49.3	359	2 T06982	glutinin low molec

30	37	49.3	374	2 T05923	glutinin low molec
31	37	49.3	395	2 B90386	hypothetical prote
32	37	49.3	403	2 B71378	probable recf prot
33	37	49.3	471	2 AC2014	two-component sens
34	37	49.3	481	2 T48404	3-deoxy-D-manno-oc
35	37	49.3	587	2 AD3519	2,3-dihydroxybenzo
36	37	49.3	636	2 T38010	hypothetical prote
37	37	49.3	891	2 G84693	hypothetical prote
38	37	49.3	1038	2 T15098	probable prolina-r
39	37	49.3	3255	2 G81702	adherence factor T
40	36	48.0	186	2 E72660	hypothetical prote
41	36	48.0	201	2 AG2376	transcription regu
42	36	48.0	244	2 A72551	hypothetical prote
43	36	48.0	311	2 S57863	pyrrolime-5-carbox
44	36	48.0	351	2 A34201	bone morphogenetic
45	36	48.0	366	2 F72703	probable heat shock

#### ALIGNMENTS

RESULT 1  
EZHU  
coagulation factor VIII precursor [validated] - human  
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant co  
C:Species: Homo sapiens (man)  
C>Date: 28-Aug-1985 #sequence, revision 28-Aug-1985 #text, change 08-Dec-2000  
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S65527; S66445;  
R:Gitschier, J.; Wood, W.I.  
Hum. Mol. Genet. 1, 199-200, 1992  
A>Title: Sequence of the exon-containing regions of the human factor VIII gene.  
A:Reference number: I54318; MUID:93265012  
A:Accession: I54318  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1921, S', 1923-2351 <RES>  
A:Cross-references: GB:M88648; NID:9182381; PIDN:AAA52420.1; PID:9182383  
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeb  
Nature 312, 330-337, 1984  
A>Title: Expression of active human factor VIII from recombinant DNA clones.  
A:Reference number: A00525; MUID:85061548  
A:Accession: A00525  
A:Molecule type: mRNA  
A:Residues: 1-2351 <MOO>  
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179  
R:Hoole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D  
s, D.N.; Hewick, R.M.  
Nature 312, 342-347, 1984  
A>Title: Molecular cloning of a cDNA encoding human antihemophilic factor.  
A:Reference number: I58059; MUID:85061550  
A:Accession: I58059  
A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-74, V', 76-1259, E', 1261-2351 <RES>  
A:Cross-references: GB:K01740; NID:9182802; PIDN:AAA52484.1; PID:9182803  
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K  
B.; Randolph, A.; Ordea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.;  
DNA 4, 333-349, 1985  
A>Title: Characterization of the polypeptide composition of human factor VIII:C and t  
A:Reference number: A23584; MUID:86081164  
A:Accession: A23584  
A:Molecule type: mRNA  
A:Residues: 1-2351 <TRU>  
A:Cross-references: GB:M14113; NID:9182817; PIDN:AAA52485.1; PID:9182818  
R:Paton, D.; Rodriguez, H.; Vohar, G.A.  
Biochemistry 25, 505-512, 1986  
A>Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag  
ity.  
A:Reference number: A26174; MUID:86159740  
A:Accession: A26174  
A:Molecule type: protein  
A:Residues: 20-36; 382-399, X', 401-402; 1668-1678; 1709-1722, D', 1723-1725; 1741-1755 <EA  
R:Pittman, D.D.; Wang, J.H.; Kautman, R.J.

Biochemistry 31, 3315-3325, 1992  
 A>Title: Identification and functional importance of tyrosine sulfate residues within re  
 A:Reference number: A42348; MUID:92207952  
 A:Accession: A42348  
 A:Molecule type: protein  
 A:Residues: 20-36;556-371,392-408;582-594,1668-1669, 'X', 1671,1672-1692,1693-1708,1709-17  
 A:Experimental source: recombinant material from Chinese hamster ovary cells  
 A>Note: sequence extracted from NCBI backbone and corrected to correspond with the publi  
 R:Pay, P.J.; Smudzin, T.M.  
 J. Biol. Chem. 264, 14005-14010, 1989  
 A>Title: Interdomain fluorescence energy transfer in human factor VIII.  
 A:Reference number: A43986; MUID:89340500  
 A:Accession: A43986  
 A:Molecule type: protein  
 A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <RAY>  
 R:Leys, A.; van Schijndel, H.B.; Niehs, C.; Hutner, W.B.; Verbeet, M.P.; Mertens, K.;  
 J. Biol. Chem. 266, 740-746, 1991  
 A>Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for  
 A:Reference number: A56109; MUID:91093266  
 A:Contents: annotation; sulfation  
 R:Gitschier, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,  
 Nature 312, 326-330, 1984  
 A>Title: Characterization of the human factor VIII gene.  
 A:Reference number: A56196; MUID:85061547  
 A:Contents: annotation; introns  
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.  
 Protein Sci. 4, 740-746, 1995  
 A>Title: Locations of disulfide bonds and free cysteines in the heavy and light chains c  
 A:Reference number: A56216; MUID:9538127  
 A:Contents: annotation; disulfide bonds  
 A>Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls  
 R:Kjake, M.; Hedling, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.  
 Eur. J. Biochem. 234, 773-779, 1995  
 A>Title: Amino acid residues 721-729 are required for full factor VIII activity.  
 A:Reference number: S63527; MUID:96163459  
 A:Accession: S63527  
 A:Molecule type: protein  
 A:Residues: 733-752;753-759 <KJ>  
 R:Rind, P.; Larsson, K.; Spita, J.; Sydow-Baechman, M.; Almstedt, A.; Gray, E.; Sandberg  
 Eur. J. Biochem. 232, 19-27, 1995  
 A>Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction  
 A:Reference number: S66445; MUID:96048024  
 A:Accession: S66445  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1668-1685 <LIN>  
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr  
 C:Genetics: GDB:F8C  
 A:Gene: F8C  
 A:Cross-references: GDB:119124; OMIM:306700  
 A:Map position: Xq28-Xq28  
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63  
 C:Function:  
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro  
 A:Pathway: blood coagulation  
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase  
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-2351/Product: coagulation factor VIII #status experimental <MAI>  
 F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>  
 F:20-356/Domain: A1 <DA1>  
 F:23-348/Domain: ferroxidase repeat homology <FO1>  
 F:392-759/Domain: A2 <DA2>  
 F:402-730/Domain: ferroxidase repeat homology <FO2>  
 F:760-1667/Domain: B <DBO>  
 F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACL>  
 F:1716-2038/Domain: A3 <DA3>  
 F:1716-2038/Domain: ferroxidase repeat homology <FO3>  
 F:2039-2191/Domain: C1 <DC1>  
 F:2039-2188/Domain: discoidin I amino-terminal homology <DN1>  
 F:2192-2345/Domain: C2 <DC2>  
 F:2192-2345/Domain: discoidin I amino-terminal homology <DN2>  
 F:60, 258, 601, 776, 803, 847, 919, 962, 982, 1020, 1024, 1074, 1085, 1204, 1274, 1278, 1301, 1319, 1403, 1

F:172-198, 267-348, 547-573, 649-730, 1851-1877, 1918-1922, 2040-2188/disulfide bonds: #sta  
 F:355-356/cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted  
 F:365-737, 738, 742, 1663, 1699/Binding site: sulfate (Tyr) (covalent) #status experiment  
 F:391-392/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment  
 F:414, 466/Binding site: sulfate (Tyr) (covalent) #status predicted  
 F:759-760/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment  
 F:1667-1668/cleavage site: Arg-Glu (unidentified proteinase) #status experimental  
 F:1708-1709/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experime  
 F:1740-1741/cleavage site: Arg-Ala (coagulation factor Xa) #status experimental  
 F:2193-2345/disulfide bonds: #status predicted

Query Match 86.7%; Score 65; DB 1; Length 2351;  
 Best Local Similarity 100.0%; Pred. No. 0.01;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1656 SGNPVLKRHR 1667

Query 2  
 A47004  
 coagulation factor VIII precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 18-Jun-1999  
 C:Accession: A47004  
 R:Eider, B.; Lachich, D.; Gitschier, J.  
 Genomics 16, 374-379, 1993  
 A>Title: Sequence of the murine factor VIII cDNA.  
 A:Reference number: A47004; MUID:93300511  
 A:Accession: A47004  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2319 <ELD>  
 A:Cross-references: GB:105573; MID:9192456; PIDN:AAA37385.1; PID:9192457  
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxid  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:23-349/Domain: ferroxidase repeat homology <FO1>  
 F:402-730/Domain: ferroxidase repeat homology <FO2>  
 F:1686-2006/Domain: ferroxidase repeat homology <FO3>  
 F:2007-2156/Domain: discoidin I amino-terminal homology <DN1>  
 F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Query Match 74.7%; Score 56; DB 2; Length 2319;  
 Best Local Similarity 91.7%; Pred. No. 0.36;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query 3 SGNPVLKRHR 14  
 Db 1629 SGNPVLKRHR 1640  
 RESULT 3  
 A25945  
 coagulation factor VIII - pig (fragment)  
 C:Species: Sus scrofa domestica (domestic pig)  
 C>Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 21-Jul-2000  
 C:Accession: A25945  
 R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murtha, P.; Wasley, L.C.; Kaufman, R.J.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986  
 A>Title: A large region (approx 195 kDa) of human factor VIII is dispensable for in v  
 A:Reference number: A25945; MUID:86287369  
 A:Accession: A25945  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-869 <TCO>  
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxid

Query Match 62.7%; Score 47; DB 2; Length 869;  
 Best Local Similarity 88.9%; Pred. No. 4.9;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 PPVLKRHR 14  
|||||

Db 737 PPVLKRHR 745

## RESULT 4

T42763  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jun-2000  
C:Accession: T42763  
R:Jollard, P.  
Submitted to the EMBL Data Library, August 1996  
A:Reference number: Z22269  
A:Accession: T42763  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2133 <L0L>  
A:Cross-references: EMBL:U49517; NID:g1511633; PIDN:AAB06705.1  
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase  
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pia  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-2133/Product: coagulation factor VIII #status predicted <MAT>  
F:23-349/Domain: ferroxidase repeat homology <FOX1>  
F:402-730/Domain: ferroxidase repeat homology <FOX2>  
F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

## Query Match

Best Local Similarity 62.7%; Score 47; DB 2; Length 2133;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 PPVLKRHR 14  
|||||

Db 1441 PPVLKRHR 1449

## RESULT 5

S70120  
ATP-dependent RNA helicase homolog YDR291W - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D9819.1  
C:Species: Saccharomyces cerevisiae  
C>Date: 24-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 02-Feb-2001  
C:Accession: S70120  
R:Patton, L.  
Submitted to the EMBL Data Library, May 1996  
A:Description: The sequence of S. cerevisiae cosmid 9819.  
A:Reference number: S70114  
A:Accession: S70120  
A:Molecule type: DNA  
A:Residues: 1-1077 <F0U>  
A:Cross-references: EMBL:U51031; NID:g1332635; PIDN:AAB64466.1; PID:g1230655; MIPS:YDR29  
C:Genetics:  
A:Map position: 4R  
A>Note: YDR291W  
C:Keywords: ATP; nucleotide binding; P-loop  
F:312-319/Region: nucleotide-binding motif A (P-loop)  
F:419-424/Region: nucleotide-binding motif B  
F:423-426/Region: DEXH motif

## Query Match

Best Local Similarity 56.0%; Score 42; DB 2; Length 1077;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 NPVLKRHR 14  
|||||

Db 500 NPVLKRHR 509

## RESULT 6

B87329  
conserved hypothetical protein CC0645 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: B87329  
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87329  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-188 <STO>  
A:Cross-references: GB:AE005673; NID:g13421858; PIDN:AAK22630.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0645

## Query Match

Best Local Similarity 54.7%; Score 41; DB 2; Length 188;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 SONPVLRHRQ 13  
::|||

Db 143 AEGPVLRRHK 153

## RESULT 7

A33733  
spectrin alpha chain - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: A33733; A49468; B49468  
R:Dubreuil, R.R.; Byers, T.J.; Sillman, A.L.; Bar-Zvi, D.; Goldstein, L.S.B.; Branton J. Cell Biol. 109, 2197-2205, 1989  
A:Title: The complete sequence of Drosophila alpha-spectrin: conservation of structure  
A:Reference number: A33733; MUID:90037215  
A:Accession: A33733  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2415 <DDB>  
A:Cross-references: GB:M26400; NID:g158488; PIDN:AA28907.1; PID:g158489  
R:Lee, J.K.; Coyne, R.S.; Dubreuil, R.R.; Goldstein, L.S.; Branton, D.  
J. Cell Biol. 123, 1797-1809, 1993  
A:Title: Cell shape and interaction defects in alpha-spectrin mutants of Drosophila m  
A:Accession: A49468; MUID:94103334  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-150 <LEF>  
A:Cross-references: GB:S67765; NID:g544666; PIDN:AAB29441.1; PID:g544667  
A>Note: sequence extracted from NCBI backbone (NCBIN:141786, NCBIN:141790, NCBIN:1417  
A:Accession: B49468  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2192-2415 <LE2>  
A:Cross-references: GB:S67765; NID:g544666; PIDN:AAB29442.1; PID:g544668  
A>Note: sequence extracted from NCBI backbone (NCBIN:141794)  
C:Genetics:  
A:Gene: FlyBase:scgr  
A:Cross-references: FlyBase:FBgn0003470  
C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectr  
C:Keywords: actin binding; cytoskeleton; EF hand  
F:46-150/Domain: spectrin/dystrophin repeat homology <SP1>  
F:151-256/Domain: spectrin/dystrophin repeat homology <SP2>  
F:257-362/Domain: spectrin/dystrophin repeat homology <SP3>  
F:363-468/Domain: spectrin/dystrophin repeat homology <SP4>  
F:469-574/Domain: spectrin/dystrophin repeat homology <SP5>  
F:575-679/Domain: spectrin/dystrophin repeat homology <SP6>  
F:680-785/Domain: spectrin/dystrophin repeat homology <SP7>  
F:786-891/Domain: spectrin/dystrophin repeat homology <SP8>

F.892-926/Domain: spectrin/dystrophin repeat homology #status atypical <SP>  
F.977-1024/Domain: SH3 homology <SH3>  
F.1072-1179/Domain: spectrin/dystrophin repeat homology <SP10>  
F.1180-1285/Domain: spectrin/dystrophin repeat homology <SP11>  
F.1286-1391/Domain: spectrin/dystrophin repeat homology <SP12>  
F.1392-1497/Domain: spectrin/dystrophin repeat homology <SP13>  
F.1498-1604/Domain: spectrin/dystrophin repeat homology <SP14>  
F.1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>  
F.1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>  
F.1817-1922/Domain: spectrin/dystrophin repeat homology <SP17>  
F.1923-2029/Domain: spectrin/dystrophin repeat homology <SP18>  
F.2038-2143/Domain: spectrin/dystrophin repeat homology <SP19>  
F.2152-2257/Domain: spectrin/dystrophin repeat homology <SP20>  
F.2265-2297/Domain: calmodulin repeat homology <EF1>  
F.2308-2340/Domain: calmodulin repeat homology <EF2>

	Query Match	Score 41;	DB 1;	Length 2415;
	Best Local	Similarity	58.3%;	
	Matches	7;	Conservative	3;
			Mismatches	2;
			Indels	0;
			Gaps	0
Q7	1	SFSQNPVILKRH	12	
		:   :   :		
Db	1995	SHAQSPAILKRH	2006	

RESULT 8  
T40224  
protein kinase - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: T40224  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Badcock, K.; Churcher, C.M.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z11914  
A:Accession: T40224  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-423 <WOO>  
A:Cross-references: EMBL:AL022172; PIDN:CAAI8163.1; GSPDB:GN00067; SPDB:SPBC32C12.03c  
A:Experimental source: strain 972h-, cosmid c32C12  
C:Genetics:  
A:Gene: SPDB:SPBC32C12.03c  
A:Map position: 2  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match	53.3%;	Score 40;	DB 2;	Length 423;
Best Local Similarity	77.8%;	Pred. No. 38;		
Matches	7;	Conservative	0;	Mismatches
			2;	Indels
				Gaps
				0

```

RESULT      9
G90430
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: G90430
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan-
joni, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragun, M.A.; Sensen, C.W.; Van der Oost, J.
Submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90430
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1506 <KUP>
A:Cross-references: GB:AE006641; NID:q13815879; PIDN:AAK42702.1; GSPDB:GN00155

```

C;Genetics:  
A;Gene: SS02575

Query Match	52.0%;	Score 39;	DB 2;	Length 506;
Best Local	Similarity 60.0%;	Pred. No. 68;		
Matches	6;	Conservative	2;	Mismatches
			2;	Indels
				Gaps
				0;

```
QY      2  FSQNPVLR  11
          ||::||:  |
Db      379  FSRNPPIFR  388
```

RESULT 10  
JN0602  
rfb protein - Escherichia coli (strain B41)  
N:Alternate names: protein B  
C:Species: Escherichia coli  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999  
C:Accession: JN0602; S33065  
R:Cheah, K.C.; Manning, P.A.  
Gene 123, 9-15, 1993  
A:Title: Inactivation of the Escherichia coli B41 (O101:K99/F41) rfb gene encoding an  
A:Reference number: JN0602; MUID:93138438  
A:Accession: JN0602  
A:Molecule type: DNA  
A:Residues: 1-597 <CHR>  
A:Cross-references: EMBL:X59852; NID:942718; PID:CAA42515.1; PID:942719  
A:Comment: This protein is located in the cytoplasmic membrane and has a role in O-ant  
C:Genetics:  
A:Gene: rfb  
C:Keywords: membrane protein

Query Match	52.0%;	Score 39;	DB 2;	Length 697;
Best Local Similarity	61.5%;	Pred. NO. 95;		
Matches	8;	Conservative	1;	Mismatches 4;
				Indels 0;
				Gaps 0;
QY	2	FSQNPVLRKHOR	14	
	1	1	1	1
	1	1	1	1
Db	684	FGCVNPVLRKHOR	696	

RESULT 11  
T00327  
polyprotein - infectious flacherie virus  
C:Species: infectious flacherie virus  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T00327  
R:Issawa, H.; Asano, S.; Sahara, K.; Iizuka, T.; Bando, H.  
Arch. Virol. 143, 127-143, 1998  
A:Title: Analysis of genetic information of an insect picorna-like virus, Infectious  
ke) viruses.  
A:Reference number: Z14139; MUID:98166871  
A:Accession: T00327  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: genomic RNA  
A:Residues: 1-3085 <ISA>  
A:Cross-references: EMBL:AB000906; NID:g3025414; PIDN:BAA25371.1; PID:g3025415  
C:Keywords: polyprotein

Query Match	52.0%	Score 39	DB 2	Length 3085
Best Local	Similarity 54.5%	Pred No. 4.4e+02		
Matches 6	Conservative	3	Mismatches 2	Indels 0
			Gaps	0

QY 2 FSQNPVLKRH 12 1:1 1: 1:1  
Db 2070 FNQNSPIWKQ 2080

RESULT	12
H81702	



adherence factor TC0438 [imported] - Chlamydia muridarum (strain Nigg)  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C:Accession: H81702  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255  
 A:Accession: H81702  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-335 <TEP>  
 A:Cross-references: GB:AE002311; GB:AE002160; NID:g7190477; PIDN:AAF39292.1; PID:g719048  
 A:Experimental source: strain Nigg (Mopn)  
 C:Genetics:  
 A:Gene: TC0438

Query Match 52.0%; Score 39; DB 2; Length 3335;  
 Best Local Similarity 87.5%; Pred. No. 4.7e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONPPVLK 10  
 |||||:  
 Db 2174 SONPPLK 2181

RESULT 13  
 B90332  
 Hypothetical protein S501708 [imported] - Sulfolobus solfataricus transposon ISCI229  
 C:Species: Sulfolobus solfataricus  
 C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 18-Jul-2001  
 C:Accession: B90332; B90341  
 R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aveyez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.O.; Medina, N.; Peng, X.; Thn-Ngoc, H.P.; Redder, F.  
 arrett, R.A.; Ragen, M.A.; Sensen, C.W.; Van der Oost, J.  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: B90332  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-131 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13814955; PIDN:AAK41913.1; GSPDB:GN00155  
 A:Accession: B90341  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-131 <KU2>  
 A:Cross-references: GB:AE006641; NID:g13815041; PIDN:AAK41985.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: S501708; S501791

Query Match 50.7%; Score 38; DB 2; Length 131;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 FSONPPVLKRHO 13  
 :|||:|:  
 Db 84 YSRNPTLKLHE 95

RESULT 14  
 S50398  
 ribosomal protein L13.e.B, cytosolic - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YW575.11c; protein YMR142c  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 22-Oct-1999  
 C:Accession: S50398  
 R:Badcock, K.; Churcher, C.  
 Submitted to the EMBL Data Library, December 1994  
 A:Reference number: S50388

A:Accession: S50398  
 A:Molecule type: DNA  
 A:Residues: 1-199 <BAD>  
 A:Cross-references: EMBL:Z47071; NID:g606429; PIDN:CAAB7356.1; PID:g606440; GSPDB:GNO  
 C:Genetics:  
 A:Gene: MIPS:YMR142c  
 A:Map position: 13R  
 A:Insertions: 2/1  
 C:Superfamily: rat ribosomal protein L13  
 C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 50.7%; Score 38; DB 1; Length 199;  
 Best Local Similarity 50.0%; Pred. No. 39;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 SFSQNPVLKRHR 14  
 :|:|:|:  
 Db 2 AISKNLPLKNHFR 15

RESULT 15  
 S67618  
 ribosomal protein L13.e.A, cytosolic - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein D2445; protein YDL082w  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 20-Jun-2000  
 C:Accession: S67618  
 R:Wambolt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
 Submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67618  
 A:Accession: S67618  
 A:Molecule type: DNA  
 A:Residues: 1-199 <RAM>  
 A:Cross-references: EMBL:Z74130; NID:g1431103; PIDN:CA98648.1; PID:g1431104; MIPS:YD  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Map position: 4L  
 A:Insertions: 2/1  
 A:Note: YDL082w  
 C:Superfamily: rat ribosomal protein L13  
 C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 50.7%; Score 38; DB 2; Length 199;  
 Best Local Similarity 50.0%; Pred. No. 39;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 SFSQNPVLKRHR 14  
 :|:|:|:  
 Db 2 AISKNLPLKNHFR 15

Search completed: June 13, 2002, 12:15:05  
 Job time: 203 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: June 13, 2002, 12:12:52 ; Search time 10.38 Seconds

(without alignments)  
52.223 Million cell updates/sec

Title: US-09-740-211-15

Perfect score: 75

Sequence: 1 SFSQNPVLKRQR 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	65	86.7	2351 1	FA8_HUMAN
2	56	74.7	2319 1	FA8_MOUSE
3	47	62.7	2133 1	FA8_PIG
4	41	54.7	2415 1	SPCA_DROME
5	40	53.3	330 1	DUSB_HUMAN
6	38	50.7	199 1	R13A_YEAST
7	38	50.7	199 1	R13B_YEAST
8	38	50.7	199 1	GDB1_WHEAT
9	38	50.7	436 1	Y177_TREPA
10	38	49.3	1333 1	SOS1_HUMAN
11	37	49.3	356 1	GLTA_WHEAT
12	36	48.0	94 1	F16B_MOUSE
13	36	48.0	311 1	PROC_NEUR
14	36	48.0	366 1	HTPX_AERPE
15	36	48.0	394 1	BMP2_MOUSE
16	36	48.0	561 1	YGG4_YEAST
17	36	48.0	737 1	OPT1_DROME
18	36	48.0	873 1	CPHA_STYX3
19	36	48.0	1522 1	MRP3_RAT
20	36	48.0	2128 1	SPCB_MOUSE
21	35.5	47.3	1833 1	KEP2_HUMAN
22	35	46.7	289 1	HTPX_PYRAB
23	35	46.7	292 1	HTPX_PYRAB
24	35	46.7	314 1	NODD_AZOCA
25	35	46.7	316 1	LYTB_ACTCA
26	35	46.7	516 1	Y067_MYCGE
27	35	46.7	529 1	DNB2_ADE02
28	35	46.7	529 1	DNB2_ADE05
29	35	46.7	533 1	LCPI_MOUSE
30	35	46.7	549 1	CET1_YEAST
31	35	46.7	614 1	ZF28_MOUSE
32	35	46.7	656 1	YAN9_SCHPO
33	35	46.7	658 1	UVRD_RICPR

## ALIGNMENTS

34	35	46.7	743 1	MYBB_XENLA	P52551 xenopus lae
35	35	46.7	758 1	PKAI_YEAST	P41909 saccharomyc
36	35	46.7	933 1	KCP3_DROME	P32023 drosophila
37	35	46.7	1014 1	UVRA_STRCO	Q92507 streptomyc
38	35	46.7	1184 1	CHSD_EMENT	P78611 emericeila
39	35	46.7	1319 1	SOS1_MOUSE	Q62245 mus musculu
40	35	46.7	1716 1	RPA1_RAT	O54889 rattus norv
41	35	46.7	4486 1	DYH9_HUMAN	O99YC9 homo sapien
42	34.5	46.0	498 1	Z271_HUMAN	Q14591 homo sapien
43	34.5	46.0	580 1	ZF35_MOUSE	P13620 mus musculu
44	34.5	46.0	711 1	Z175_HUMAN	O94473 homo sapien
45	34	45.3	105 1	RL24_THEMA	P38513 thermotoga

RESULT 1	FA8_HUMAN	STANDARD:	PRT: 2351 AA.
AC	P00451:		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Coagulation factor VIII precursor (Procoagulant component)		
DE	(Antihemophilic factor) (AHF).		
GN	P8 OR PRC.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=86081164; PubMed=3935400;		
RA	Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,		
RA	Harlog K., Kuo C.H., Maslitz F.R., Merryweather J.P., Najjarian R.,		
RA	Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,		
RA	Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,		
RA	Nordfang O., Ezban M.;		
RT	"Characterization of the polypeptide composition of human factor		
RT	VIII:C and the nucleotide sequence and expression of the human kidney		
RT	cDNA.";		
RL	DNA 4:333-349(1985).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=85061548; PubMed=6438526;		
RA	Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,		
RA	Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,		
RA	Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;		
RT	"Expression of active human factor VIII from recombinant DNA clones.";		
RL	Nature 312:330-337(1984).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=85061550; PubMed=6438528;		
RA	Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,		
RA	Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,		
RA	Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,		
RA	Hewick R.M.;		
RT	"Molecular cloning of a cDNA encoding human antihemophilic factor.";		
RL	Nature 312:342-347(1984).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=93265012; PubMed=1303178;		
RA	Gitschier J., Wood W.I.;		
RT	"Sequence of the exon-containing regions of the human factor VIII		
RT	gene.";		
RL	Hum. Mol. Genet. 1:199-200(1992).		
RN	[5]		
RP	SEQUENCE OF 2064-2070 FROM N.A.		
RA	de Water N.S., Williams R., Birowett P.J.;		
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.		
RN	[6]		
RP	SULFATION OF TYR-1699.		

- RX MEDLINE-91093266; PubMed-1898735;  
 RA Leyte A., van Schijndel H.B., Niehs C., Hutner W.B., Verbeet M.P.,  
 RA Mertens K., van Mourik J.A.;  
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is  
 RT essential for the interaction of factor VIII with von Willebrand  
 RT factor.";  
 RL J. Biol. Chem. 266:740-746(1991).  
 RN [17]  
 RN SULFATION.  
 RP MEDLINE-92207952; PubMed-1554716;  
 RA Pittman D.D., Wang J.H., Kaufman R.J.;  
 RT "Identification and functional importance of tyrosine sulfate  
 RT residues within recombinant factor VIII.";  
 RL Biochemistry 31:3315-3325(1992).  
 RN [18]  
 RN STRUCTURE BY NMR OF 2322-2343.  
 RP MEDLINE-95200924; PubMed-7893714;  
 RA Gilbert G.E., Baleja J.D.;  
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an  
 RT amphipathic structure as determined by NMR spectroscopy.";  
 RL Biochemistry 34:3022-3031(1995).  
 RN [19]  
 RN REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.  
 RP MEDLINE-91221499; PubMed-1902642;  
 RA Gitschler J.;  
 RT "The molecular basis of hemophilia A.";  
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).  
 RN [10]  
 RN REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.  
 RP MEDLINE-89088506; PubMed-2491949;  
 RA White G.C. II, Shoemaker C.B.;  
 RT "Factor VIII gene and hemophilia A.";  
 RL Blood 73:1-12(1989).  
 RN [11]  
 RN REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.  
 RP MEDLINE-95245332; PubMed-7728145;  
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;  
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";  
 RL Hum. Mutat. 5:1-22(1995).  
 RN [12]  
 RN VARIANT GLN-2326.  
 RP MEDLINE-86235434; PubMed-3012775;  
 RA Gitschler J., Wood W.I., Shuman M.A., Lawn R.M.;  
 RT "Identification of a missense mutation in the factor VIII gene of a  
 RT mild hemophiliac.";  
 RL Science 232:1415-1416(1986).  
 RN [13]  
 RN VARIANT PRO-2135.  
 RP MEDLINE-88096539; PubMed-3122181;  
 RA Levinson B., Jancso R.L., Phillips J.A. III, Gitschler J.;  
 RT "A novel missense mutation in the factor VIII gene identified by  
 RT analysis of amplified hemophilia DNA sequences.";  
 RL Nucleic Acids Res. 15:9797-9805(1987).  
 RN [14]  
 RN VARIANT GLN-2228.  
 RP MEDLINE-88191889; PubMed-2833855;  
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,  
 RA Kazazian H.H.;  
 RT "Nonsense and missense mutations in hemophilia A: estimate of the  
 RT relative mutation rate at CG dinucleotides.";  
 RL Am. J. Hum. Genet. 42:718-725(1988).  
 RN [15]  
 RN VARIANT GLY-291.  
 RP MEDLINE-88220354; PubMed-2835904;  
 RA Youssoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,  
 RA Antonarakis S.E.;  
 RT "Moderately severe hemophilia A resulting from Glu->Gly substitution  
 RT in exon 7 of the factor VIII gene.";  
 RL Am. J. Hum. Genet. 42:867-871(1988).  
 RN [16]  
 RN VARIANT CYS-1708.  
 RP MEDLINE-89274393; PubMed-2499363;  
 RA O'Brien D.P., Tuddenham E.G.;  
 RT "Purification and characterization of factor VIII 1,689-Cys: a  
 RT nonfunctional cofactor occurring in a patient with severe hemophilia  
 RT A.";  
 RL Blood 73:2117-2122(1989).  
 RN [17]  
 RN VARIANT CYS-391.  
 RP MEDLINE-90001543; PubMed-2506948;  
 RA Shima M., Ware J., Yoshioke A., Fukui H., Pulcher C.A.;  
 RT "An arginine to cysteine amino acid substitution at a critical  
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";  
 RL Blood 74:1612-1617(1989).  
 RN [18]  
 RN VARIANT LEU-189.  
 RP MEDLINE-90057680; PubMed-2510835;  
 RA Chan V., Chan T.K., Tong T.M., Todd D.;  
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene  
 RT resulting in moderately severe hemophilia A.";  
 RL Blood 74:2688-2691(1989).  
 RN [19]  
 RN VARIANT LEU-2326.  
 RP MEDLINE-89197216; PubMed-2495245;  
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;  
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26  
 RT of the factor VIII gene.";  
 RL Hum. Genet. 81:335-338(1989).  
 RN [20]  
 RN VARIANT HIS-391.  
 RP MEDLINE-89264602; PubMed-2498882;  
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,  
 RA Fujimaki M., Hoyer L.W.;  
 RT "Direct characterization of factor VIII in plasma: detection of a  
 RT mutation altering a thrombin cleavage site  
 RT (arginine-372-->histidine).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).  
 RN [21]  
 RN VARIANT CYS-1708.  
 RP MEDLINE-90105723; PubMed-2104766;  
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,  
 RA Phillips J.A. III, Jancso R.L., Hoyer L.W.;  
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to  
 RT Cys) in the factor VIII gene of two unrelated patients with  
 RT cross-reacting material-positive hemophilia A.";  
 RL Blood 75:384-389(1990).  
 RN [22]  
 RN VARIANTS GLN-2228 AND LEU-2326.  
 RP MEDLINE-90123183; PubMed-2105106;  
 RA Casula U., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,  
 RA Mancuso G., Morfini M., de Biasi R., Baudou F., Cardonara A.;  
 RT "Recurrent mutations and three novel rearrangements in the factor  
 RT VIII gene of hemophilia A patients of Italian descent.";  
 RL Blood 75:662-670(1990).  
 RN [23]  
 RN VARIANT CYS-391.  
 RP MEDLINE-90329422; PubMed-1973901;  
 RA Patinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;  
 RT "CRM+ hemophilia A due to a missense mutation (372->Cys) at the  
 RT internal heavy chain thrombin cleavage site.";  
 RL Br. J. Haematol. 75:73-77(1990).  
 RN [24]  
 RN VARIANTS PHE-1699 AND CYS-1708.  
 RP MEDLINE-90152691; PubMed-2105906;  
 RA Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,  
 RA Kazazian H.H., Antonarakis S.E.;  
 RT "Characterization of mutations in the factor VIII gene by direct  
 RT sequencing of amplified genomic DNA.";  
 RL Genomics 6:65-71(1990).  
 RN [25]  
 RN VARIANTS CYS-1728 AND ASP-1941.  
 RP MEDLINE-90169988; PubMed-2106480;  
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,  
 RA Kazazian H.H.;  
 RT "Use of denaturing gradient gel electrophoresis to detect point  
 RT mutations in the factor VIII gene."

Query Match 86.7%; Score 65; DB 1; Length 2351;  
 Best Local Similarity 100.0%; Pred. No. 0.0038;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 SONPPVLRHOR 14  
 |||||||||  
 Db 1656 SONPPVLRHOR 1667

RESULT 2  
 F8\_MOUSE STANDARD; PRT; 2319 AA.

AC Q06194; 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Coagulation factor VIII precursor (Procoagulant component).  
 GN CF8 OR F8 OR F8C.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;  
 RA MEDLINE=93300511; PubMed=8314577;  
 RX Elder B., Lakich D., Gitschler J.;  
 RT "Sequence of the murine factor VIII cDNA."  
 RL Genomics 16:374-379(1993).  
 CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS  
 AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE  
 ACTIVATED FORM, FACTOR XA.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.  
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF  
 2 PLASTOCYANIN-LIKE REPEATS.  
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.  
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: I05573; AAA37385.1; -.  
 CC PIR: A47004; A47004.  
 DR HSSP; P00451; ICRG.  
 DR MGD; MGI:88383; F8.  
 DR InterPro; IPR001117; Cu-oxidase.  
 DR InterPro; IPR000421; F5/8\_C.  
 DR Pfam; PF00394; Cu-oxidase; 3.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 2.  
 DR SMART; SM00231; F5/8C\_2.  
 DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 3.  
 DR PROSITE; PS01285; F5/8C\_1; 2.  
 DR PROSITE; PS01286; F5/8C\_2; 2.  
 KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;  
 KW Signal; Glycoprotein; Sulfation.  
 KM SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 2319 COAGULATION FACTOR VIII.  
 FT DOMAIN 20 349 F5/8 TYPE A 1.  
 FT DOMAIN 207 349 PLASTOCYANIN-LIKE 1.  
 FT DOMAIN 207 349 PLASTOCYANIN-LIKE 2.  
 FT DOMAIN 399 573 F5/8 TYPE A 2.  
 FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.  
 FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.  
 FT DOMAIN 760 1640 B.  
 FT DOMAIN 1683 2008 F5/8 TYPE A 3.  
 FT DOMAIN 1683 1845 PLASTOCYANIN-LIKE 5.

FT DOMAIN 1855 2008 PLASTOCYANIN-LIKE 6.  
 FT DOMAIN 2009 2156 F5/8 TYPE C 1.  
 FT DOMAIN 2161 2313 F5/8 TYPE C 2.  
 FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
 FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
 FT SITE 1678 1679 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
 FT SITE 1324 1325 CLEAVAGE (ACTIVATION) (BY SIMILARITY).  
 FT SITE 1640 1641 CLEAVAGE (ACTIVATION) (BY SIMILARITY).  
 FT SITE 367 367 SULFATION (BY SIMILARITY).  
 FT MOD\_RES 737 737 SULFATION (BY SIMILARITY).  
 FT MOD\_RES 738 738 SULFATION (BY SIMILARITY).  
 FT MOD\_RES 742 742 SULFATION (BY SIMILARITY).  
 FT MOD\_RES 1669 1669 SULFATION (REQUIRED FOR VWF BINDING) (BY SIMILARITY).  
 FT MOD\_RES 1687 1687 SULFATION (BY SIMILARITY).  
 FT DISULFID 173 199 PROBABLE.  
 FT DISULFID 547 573 PROBABLE.  
 FT DISULFID 1819 1845 PROBABLE.  
 FT DISULFID 2008 2156 BY SIMILARITY.  
 FT DISULFID 2161 2313 BY SIMILARITY.  
 FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 601 601 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 880 880 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 958 958 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1015 1015 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1087 1087 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1136 1136 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1161 1161 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1192 1192 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1255 1255 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1268 1268 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1273 1273 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1316 1316 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1340 1340 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1797 1797 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 2105 2105 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 2319 AA; 266148 MW; PD054DE051DB2A01 CMC64;

Query Match 74.7%; Score 56; DB 1; Length 2319;  
 Best Local Similarity 91.7%; Pred. No. 0.14;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 SONPPVLRHOR 14  
 |||||||||  
 Db 1629 SQPPVLRHOR 1640

RESULT 3  
 F8\_MOUSE STANDARD; PRT; 2133 AA.

AC P12263; 095243; 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coagulation factor VIII precursor (Procoagulant component).  
 GN CF8.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fodler C., Gabriellian A.E., Garg N.S., Galt J.H., Glaser K.,  
 RA Glaser K., Garg N.S., Galt J.H., Glaser K., Glaser K.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., Moleod M.P., Moberg D.,  
 RA Merkulov G., Mishina N.V., Moberg D., Morris J., Moshnelli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RN SEQUENCE OF 1-150 AND 2192-2415 FROM N.A. AND CHARACTERIZATION.  
 RP MEDLINE=9410334; PubMed=8276898;  
 RA Lee J.K., Coyne R.S., Dubreuil R.R., Goldstein L.S.B., Branton D.;  
 RT "Cell shape and interaction defects in alpha-spectrin mutants of  
 RT *Drosophila melanogaster*.";  
 RL J. Cell Biol. 123:1797-1809(1993).  
 RN [4]  
 RN CHARACTERIZATION.  
 RP MEDLINE=88059242; PubMed=3680372;  
 RA Dubreuil R., Byers T.J., Branton D., Goldstein L.S.B., Kiehart D.P.;  
 RT "Drosophila spectrin. I. Characterization of the purified protein.";  
 RL J. Cell Biol. 105:2093-2102(1987).  
 RN [5]  
 RN EMBRYONIC LOCALIZATION.  
 RP MEDLINE=89234159; PubMed=2497103;  
 RA Pesacreta T.C., Byers T.J., Dubreuil R., Kiehart D.P., Branton D.;  
 RT "Drosophila spectrin: the membrane skeleton during embryogenesis.";  
 RL J. Cell Biol. 108:1697-1709(1989).  
 RN [6]  
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 1391-1497.  
 RP MEDLINE=94090340; PubMed=8266097;  
 RA Yan Y., Whinograd E., Vriel A., Cronin T., Harrison S.C., Branton D.;  
 RT "Crystal structure of the repetitive segments of spectrin.";  
 RL Science 262:2027-2030(1993).  
 RL -1- FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL  
 CC NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES  
 CC WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF  
 CC THE ERYTHROCYTE PLASMA MEMBRANE. ESSENTIAL FOR LARVAL SURVIVAL AND  
 CC DEVELOPMENT. STABILIZES CELL TO CELL INTERACTIONS THAT ARE  
 CC CRITICAL FOR THE MAINTENANCE OF CELL SHAPE AND SUBCELLULAR  
 CC ORGANIZATION WITHIN EMBRYONIC TISSUES.  
 CC -1- SUBUNIT: NATIVE SPECTRIN MOLECULE IS A TETRAMER COMPOSED OF TWO  
 CC ANTIPARALLEL HETERODIMERS JOINED HEAD TO HEAD SO THAT EACH END OF  
 CC THE NATIVE MOLECULE INCLUDES THE C-TERMINUS OF THE ALPHA SUBUNIT  
 CC AND THE N-TERMINUS OF THE BETA SUBUNIT. INTERACTS WITH CALMODULIN  
 CC IN A CALCIUM-DEPENDENT MANNER AND INTERACTS WITH F-ACTIN.  
 CC -1- SUBCELLULAR LOCATION: NEAR THE INNER SURFACE OF THE PLASMA  
 CC MEMBRANE OF NEARLY ALL CELLS.  
 CC -1- TISSUE SPECIFICITY: A SUBSTANTIAL POOL OF MATERNAL PROTEIN IN THE  
 CC EGG UNDERGOES DYNAMIC CHANGES IN DISTRIBUTION EARLY IN  
 CC EMBRYOGENESIS. IN GASTRULATED EMBRYO, THE HIGHEST LEVEL OF PROTEIN  
 CC IS FOUND IN THE RESPIRATORY TRACT CELLS AND THE LOWEST IN PARTS OF  
 CC THE FORMING GUT.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.  
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M26400; AAA8907.1; -  
 CC EMBL; AE003472; AAF47569.1; -  
 CC EMBL; S67762; AAB29441.2; -  
 CC EMBL; S67765; AAB29442.1; -  
 CC PIR; A33733; A33733.  
 CC PDB; 2SPC; 31-MAY-94.  
 CC Flybase; FBgn003470; Alpha-Spec.  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR001452; SH3.  
 CC InterPro; IPR002017; Spectrin.  
 CC Pfam; PF00036; ehand; 2.  
 CC Pfam; PF00018; SH3; 1.  
 CC Pfam; PF00435; spectrin; 22.  
 CC PRINTS; PR00452; SH3DOMAIN.  
 CC SMART; SM00054; EFh; 2.  
 CC SMART; SM00326; SH3; 1.  
 CC SMART; SM00150; SPEC; 20.  
 CC PROSITE; PS00018; EF\_HAND; 2.  
 CC PROSITE; PS00002; SH3; 1.  
 CC Cytochrome; Membrane; Erythrocyte; Repeat; Actin-binding;  
 CC Capturing protein; Calcium-binding; SH3 domain; 3D-structure;  
 CC Cell shape; Calmodulin-binding.  
 FT REPEAT 18 122 SPECTRIN 1.  
 FT REPEAT 123 228 SPECTRIN 2.  
 FT REPEAT 229 334 SPECTRIN 3.  
 FT REPEAT 335 440 SPECTRIN 4.  
 FT REPEAT 441 546 SPECTRIN 5.  
 FT REPEAT 547 651 SPECTRIN 6.  
 FT REPEAT 652 757 SPECTRIN 7.  
 FT REPEAT 758 863 SPECTRIN 8.  
 FT REPEAT 864 969 SPECTRIN 9.  
 FT REPEAT 970 1043 SPECTRIN 10.  
 FT REPEAT 1044 1151 SPECTRIN 11.  
 FT REPEAT 1152 1257 SPECTRIN 12.  
 FT REPEAT 1258 1363 SPECTRIN 13.  
 FT REPEAT 1364 1469 SPECTRIN 14.  
 FT REPEAT 1470 1576 SPECTRIN 15.  
 FT REPEAT 1577 1682 SPECTRIN 16.  
 FT REPEAT 1683 1788 SPECTRIN 17.  
 FT REPEAT 1789 1894 SPECTRIN 18.  
 FT REPEAT 1895 2001 SPECTRIN 19.  
 FT REPEAT 2002 2115 SPECTRIN 20.  
 FT REPEAT 2116 2229 SPECTRIN 21.  
 FT REPEAT 2230 2335 SPECTRIN 22.  
 FT CA BIND 2278 2289 EF-HAND 1 (POTENTIAL).  
 FT CA BIND 2321 2332 EF-HAND 2 (POTENTIAL).  
 FT DOMAIN 970 1029 SH3.  
 FT CONFLICT 110 110 Q -> D (IN REF. 3).  
 FT CONFLICT 1668 1668 Q -> R (IN REF. 1).  
 FT SEQUENCE 2415 AA; 278301 MW; F1F72FB995B0A37 CRC64;  
 SQ  
 Query Match 54.7%; Score 41; DB 1; Length 2415;  
 Best local Similarity 58.3%; Pred. No. 60;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 SFSPNPVLKRR 12  
 DB 1995 SHSPAILKRR 2006  
 RESULT 5





DR SGD: S0004750; RPL13B.  
 DR InterPro: IPR001380; Ribosomal\_L13E.  
 DR Pfam: PF01294; Ribosomal\_L13e; 1.  
 DR ProDom: PD004443; Ribosomal\_L13E; 1.  
 DR PROSITE: PS01104; RIBOSOMAL\_L13E; 1.  
 KW Ribosomal protein; Multigene family.  
 SQ SEQUENCE 199 AA; 22525 MW; BEA3B423EE76EF23 CRC64;

Query Match 50.7%; Score 38; DB 1; Length 199;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 SFSQNPVLRKHOR 14  
 DB 2 AISKNPILKHNFR 15

RESULT 8  
 GDB1\_WHEAT STANDARD; PRT; 304 AA.  
 AC P04729;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Gamma-gliadin B-1 precursor.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4563;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85234522; PubMed=2989281;  
 RA Okita T.W., Cheesbrough V., Reeves C.D.;  
 RT "Evolution and heterogeneity of the alpha-/beta-type and gamma-type  
 gliadin DNA sequences";  
 RL J. Biol. Chem. 260:8203-8213(1985).  
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.  
 CC -1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY  
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO  
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR  
 CC NEAR DIRECT REPEATS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: M11077; AAA34285.1; -.  
 DR InterPro: IPR003612; AAI.  
 DR InterPro: IPR001968; Cereal\_tryp\_amy1\_inh.  
 DR InterPro: IPR001954; Glla\_glu1en1n.  
 DR Pfam: PF00234; tryp\_alpha\_amy1; 1.  
 DR PRINTS: PR00208; GLIADGLUTEN.  
 DR SMART: SM00499; AAI; 1.  
 KW Seed storage protein; Repeat; Signal; Multigene family.  
 FT SIGNAL 1 23  
 FT CHAIN 24 304 GAMMA-GLIADIN B-1.  
 SO SEQUENCE 304 AA; 34252 MW; 807EBFA47A59D6D5 CRC64;

Query Match 50.7%; Score 38; DB 1; Length 304;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 SFSQNPVLRKHOR 14  
 DB 43 SFSQNPVLRKHOR 56

RESULT 9  
 Y177\_TREPA STANDARD; PRT; 436 AA.  
 ID Y177\_TREPA  
 AC 083207;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein TP0177.  
 GN TP0177.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLO;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 agent";  
 RL Science 281:375-388(1998).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AEO01201; AAC65167.1; -.  
 DR TIGR: TP0177;  
 DR Hypothetical protein; Complete proteome.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 436 AA; 49532 MW; 50A9DC1D838057AB CRC64;

Query Match 50.7%; Score 38; DB 1; Length 436;  
 Best Local Similarity 35.7%; Pred. No. 31;  
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 SFSQNPVLRKHOR 14  
 DB 351 TFSQNPVLRKHOR 364

RESULT 10  
 S0S1\_HUMAN STANDARD; PRT; 1333 AA.  
 ID S0S1\_HUMAN  
 AC 007889;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Son of sevenless protein homolog 1 (SOS-1).  
 GN S0S1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93262494; PubMed=8493579;  
 RA Chardin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,  
 RA Bar-Sagi D.;  
 RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds  
 to GRB2";  
 RL Science 260:1338-1343(1993).

[2]  
 RN STRUCTURE BY NMR OF 422-551.  
 RE MEDLINE=96043737; PubMed=9374522;  
 RA Zheng J., Chen R.H., Corbhan-Garcia S., Cahill S.M., Bar-Sagi D.,  
 RA Cowburn D.;  
 RT "The solution structure of the pleckstrin homology domain of human  
 RT SOS1. A possible structural role for the sequential association of  
 RT diffuse B cell lymphoma and pleckstrin homology domains.";  
 RL J. Biol. Chem. 272:30340-30344(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 198-551.  
 RA MEDLINE=99005193; PubMed=9790532;  
 RA Soisson S.M., Minnaal A.S., Uy M., Bar-Sagi D., Kuriyan J.;  
 RT "Crystal structure of the Dbl and pleckstrin homology domains from  
 RT the human Son of sevenless protein.";  
 RL Cell 95:259-268(1998).  
 CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
 CC -1- SUBUNIT: INTERACTS WITH GRB2.  
 CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L13857; AAA5913.1; -  
 DR PDB: 1AME; 25-FEB-98.  
 DR PDB: 1DBH; 23-DEC-98.  
 DR MIM: 182530; -  
 DR InterPro: IPR002119; Histone\_H2A.  
 DR InterPro: IPR00166; Histone\_core.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000651; RasGEFN.  
 DR InterPro: IPR001895; RasGR\_CCC25.  
 DR InterPro: IPR00219; RhoGEF.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00617; RasGEF; 1.  
 DR Pfam: PF00618; RasGEFN; 1.  
 DR Pfam: PF00621; RhoGEF; 1.  
 DR SMART: SM00414; H2A; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00147; RasGEF; 1.  
 DR SMART: SM00229; RasGEFN; 1.  
 DR SMART: SM00325; RhoGEF; 1.  
 DR PROSITE: PS00741; DH.1; FALSE\_NEG.  
 DR PROSITE: PS50010; DH.2; 1.  
 DR PROSITE: PS00720; GDS\_CDC25; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 KW Guanine-nucleotide releasing factor; 3D-structure.  
 FT DOMAIN 200 390 DH.  
 FT DOMAIN 444 548 PH.  
 FT DOMAIN 777 963 RASGEF.  
 FT DOMAIN 1258 1261 POLY-PRO.  
 SO SEQUENCE 1333 AA; 152463 MW; C6B89CCAl1A8DE45 CRC64;

Query Match 50.7%; Score 38; DB 1; Length 1333;  
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 NPVLKRH 12  
 II II II II II  
 DB 944 NPVLKRH 951

RESULT 11  
 GLTA\_WHEAT  
 ID GLTA\_WHEAT STANDARD; PRT; 356 AA.

AC P10385;  
 DT 01-MAR-1989 (Rel. 10, Last Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glutenin, low molecular weight subunit precursor.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Triticum.  
 NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=CV. YAMHILL;  
 RX MEDLINE=89083577; PubMed=3205747;  
 RA Pitts E.G., Rafalski J.A., Hedgcock C.;  
 RT "Nucleotide sequence and encoded amino acid sequence of a genomic  
 RT gene region for a low molecular weight glutenin.";  
 RL Nucleic Acids Res. 16:11376-11376(1988).  
 CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE  
 CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE  
 CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.  
 CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.  
 CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE  
 CC GROUP 1 CHROMOSOMES OF WHEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07747; CAA30570.1; -  
 DR PIR: S01992; S01992.  
 DR InterPro: IPR003612; AAI.  
 DR InterPro: IPR001768; Cereal\_1ryp\_amy1\_inh.  
 DR InterPro: IPR001954; Gli1-glutenin.  
 DR Pfam: PF00234; 1ryp\_alpha\_amy1; 1.  
 DR PRINTS: PR00208; GLIADGLUTEN.  
 DR SMART: SM00499; AAI; 1.  
 KW Seed storage protein; Repeat: Multigene family; Signal.  
 FT SIGNAL 20 356 POTENTIAL.  
 FT CHAIN 1 19 GLUTENIN, LOW MOLECULAR WEIGHT SUBUNIT.  
 SO SEQUENCE 356 AA; 41020 MW; AA2613RCDD4DCAS CRC64;

Query Match 49.3%; Score 37; DB 1; Length 356;  
 Best Local Similarity 50.0%; Pred. No. 37;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 ESONPVLKRH 13  
 III III III  
 DB 71 ESONPVLKRH 82

RESULT 12  
 F16B\_MOUSE  
 ID F16B\_MOUSE STANDARD; PRT; 94 AA.  
 AC P97323;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fructose-1,6-bisphosphatase isozyme 3 (EC 3.1.3.11) (D-fructose-1,6-  
 DE bisphosphate 1-phosphohydrolase) (FBPase) (Fragment).  
 DE FBP3.  
 GN FBP3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J;

```

RX MEDLINE=97260614; PubMed=9106734;
RA Claix J.F., Beaulieu E., Hevor T.K.;
RT "Various fructose-1,6-bisphosphatase mRNAs in mouse brain, liver,
RT kidney and heart.";
RL NeuroReport 8:617-622(1997).
CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate + H2O = D-
CC fructose 6-phosphate + phosphate.
CC -1- PATHWAY: NECESSARY FOR, AND SPECIFIC TO, GLUCONEOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE PBPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11067; CAA71946.1; -.
DR HSSP; P09467; 1PTA.
DR MGD; MGI:1330236; Fbp3.
DR InterPro; IPR000146; In_FB_phphatase.
DR Pfam; PF00316; FBPase; 1.
DR ProDom; PD001491; In_FB_phphatase; 1.
DR PROSITE; PS00124; FBPASE; PARTIAL.
RW Hydrolyase; Carbohydrate metabolism; Gluconeogenesis.
FT NON_TER 1 94
FT SEQUENCE 94 AA; 10459 MW; CB8987281DB29944 CRC64;

Query Match 48.0%; Score 36; DB 1; Length 94;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 NPPVYKRRH 14
DB 68 NPALLEYHQR 77

RESULT 13
PROC_NEUCR STANDARD; PRT; 311 AA.
AC Q12641;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN pro-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=96004701; PubMed=7565596;
RA Davis C.R., McPeck M.A., McClung C.R.;
RT "Molecular characterization of the proline-1 (pro-1) locus of
RT Neurospora crassa, which encodes delta 1-pyrroline-5-carboxylate
RT reductase.";
RL Mol. Genet. 248:341-350(1995).
CC -1- CATALYTIC ACTIVITY: L-proline + NAD(P)H.
CC -1- PATHWAY: TERMINAL (THIRD) STEP IN PROLINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE PYROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03017; AAA83568.1; -.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; 1.
RW Oxidoreductase; Proline biosynthesis; NADP.
FT SEQUENCE 311 AA; 32151 MW; 9A6D525D845984DF CRC64;

Query Match 48.0%; Score 36; DB 1; Length 311;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 NPPVYKRRH 12
DB 26 NPPALPRH 33

RESULT 14
HTPX_AERPE STANDARD; PRT; 366 AA.
AC Q9YD67;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Probable protease htpx homolog (EC 3.4.24.-).
GN HTPX OR APEL045.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankel A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Koshida N., Oguchi A., Aoki K.-I., Kudota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -1- COFACTOR: Zinc (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000060; BAA80030.1; -.
DR MEROPS; M48.004; -.
DR InterPro; IPR001915; Peptidase_M48.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF01435; Peptidase_M48; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
RW Transmembrane; Hydrolyase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 4 24
FT TRANSMEM 50 70
FT TRANSMEM 79 99
FT TRANSMEM 193 213
FT TRANSMEM 229 249
FT METAL 181 181
FT ACT_SITE 182 182
FT METAL 185 185
FT SEQUENCE 366 AA; 39742 MW; 3DE2AP596CC76584 CRC64;

```

Query Match 48.0%; Score 36; DB 1; Length 366;  
 Best Local Similarity 70.0%; Pred. No. 57;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 F50NPPVLRK 11  
 11:111111  
 DB 340 F5THPPVSKR 349

RESULT 15  
 BMP2\_MOUSE STANDARD: PRT; 394 AA.  
 ID BMP2\_MOUSE  
 AC P21274;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Bone morphogenetic protein 2 precursor (BMP-2) (BMP-2A).  
 GN BMP2 OR BMP-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94289485; PubMed=8018727;  
 RA Feng J.O., Harris M.A., Ghosh-Choudhury N., Feng M., Mundy G.R.,  
 RA Harris S.E.;  
 RT "Structure and sequence of mouse bone morphogenetic protein-2 gene  
 (BMP-2): comparison of the structures and promoter regions of BMP-2  
 and BMP-4 genes."  
 RL Biochim. Biophys. Acta 1218:221-224(1994).  
 RN [2]  
 RP SEQUENCE OF 1-351 FROM N.A.  
 RX MEDLINE=90228966; PubMed=1970330;  
 RA Dickinson M.E., Kobrin M.S., Silan C.M., Kingsley D.M., Justice M.J.,  
 RA Miller D.A., Cecil J.D., Lock L.F., Lee A., Buchberg A.M.,  
 RA Stracusa L.D., Lyons K.M., Derynck R., Hogan B.L.M., Copeland N.G.,  
 RA Jenkins N.A.;  
 RT "Chromosomal localization of seven members of the murine TGF-beta  
 superfamily suggests close linkage to several morphogenetic mutant  
 loci."  
 RL Genomics 6:505-520(1990).  
 CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.  
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L25602; AAB05665.1; -  
 DR PIR: A34201; A34201.  
 DR HSSP: P12643; 3BMP.  
 DR MCD: MGI:88177; BMP2.  
 DR InterPro: IPR001839; TGF-beta.  
 DR InterPro: IPR001111; TGF-beta.  
 DR Pfam: PF00019; TGF-beta; 1.  
 DR Pfam: PF00688; TGF-beta; 1.  
 DR ProDom: PD000357; TGF-beta; 1.  
 DR SMART: SM00204; TGF-beta; 1.  
 DR SMART: SM00250; TGF-beta; 1.  
 DR Signal: PS00250; TGF-beta; 1.  
 DR Signal: Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.  
 FT PROPEP 1 19 POTENTIAL.  
 FT SIGNAL 20 280  
 FT CHAIN 281 394 BONE MORPHOGENETIC PROTEIN 2.  
 FT DISULFID 294 359 BY SIMILARITY.  
 FT DISULFID 323 391 BY SIMILARITY.

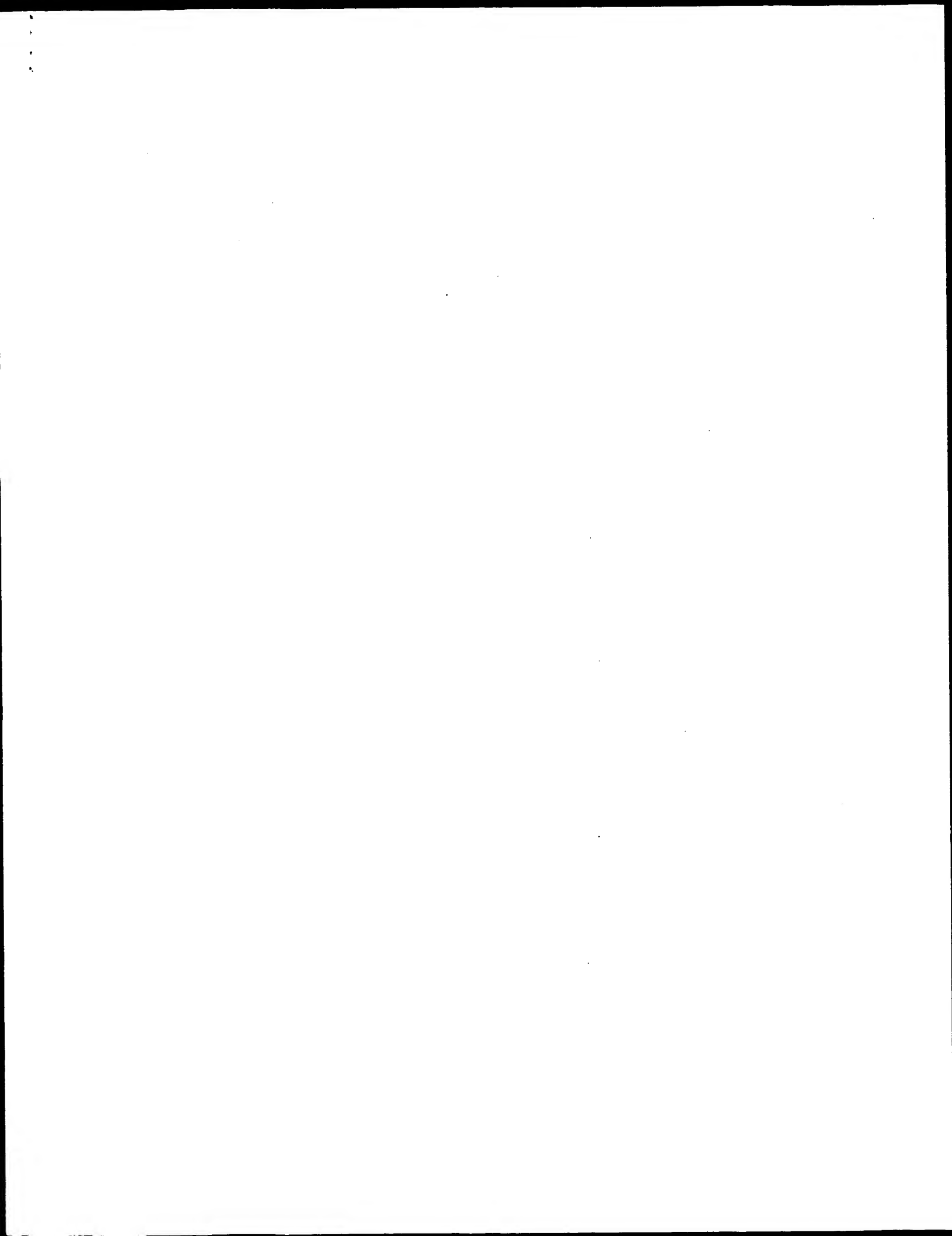
FT DISULFID 327 393 BY SIMILARITY.  
 FT DISULFID 358 358 INTERCHAIN (BY SIMILARITY).  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 110 110 T -> S (IN REF. 2).  
 FT CONFLICT 113 114 OL -> HE (IN REF. 2).  
 FT CONFLICT 271 271 G -> R (IN REF. 2).  
 FT CONFLICT 271 271  
 SQ SEQUENCE 394 AA; 44514 MW; FD6A0F10587ED54 CRC64;

Query Match 48.0%; Score 36; DB 1; Length 394;  
 Best Local Similarity 63.6%; Pred. No. 62;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 4 ONPPVLRKHOR 14  
 11:111111  
 DB 233 ENPGVSKRHVR 243

Search completed: June 13, 2002, 12:16:13  
 Job time: 201 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2002, 12:12:17 ; Search time 25 Seconds  
(without alignments)  
96.877 Million cell updates/sec

Title: US-09-740-211-15

Perfect score: 75

Sequence: 1 SFSQNPVLRKHQR 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp Unclassified:\*  
15: sp\_rvirts:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	72.0	2343	6	062730
2	54	72.0	2343	6	018806
3	42	56.0	1077	3	005549
4	41	54.7	188	16	09AA6
5	41	54.7	450	10	09FNM6
6	40	53.3	330	4	09BME3
7	40	53.3	330	4	075319
8	40	53.3	423	3	059697
9	39.5	52.7	561	4	096BVO
10	39	52.0	225	5	0903VA
11	39	52.0	291	10	09FIW3
12	39	52.0	445	17	09W0F6
13	39	52.0	506	17	097VP1
14	39	52.0	563	10	0944G0
15	39	52.0	608	10	09FHY4
16	39	52.0	697	2	005161

17	39	52.0	3085	12	070710	070710 infectious
18	39	52.0	3335	16	09PKM7	09PKM7 chlamydia m
19	38	50.7	131	17	097TX3	097TX3 sulfolobus
20	38	50.7	217	11	09D923	09D923 mus musculu
21	38	50.7	269	1	074066	074066 cenarchaeum
22	38	50.7	273	1	074043	074043 cenarchaeum
23	38	50.7	303	10	P93792	P93792 tritlicum ae
24	38	50.7	332	16	09HVG4	09HVG4 pseudomonas
25	38	50.7	348	10	042956	042956 nicotiana t
26	38	50.7	604	3	013845	013845 schizosacch
27	38	50.7	623	5	018013	018013 caenorhabdi
28	38	50.7	636	2	09F620	09F620 treponema h
29	38	50.7	1144	5	09XW61	09XW61 caenorhabdi
30	38	50.7	1217	5	P91094	P91094 caenorhabdi
31	38	50.7	1236	4	09C012	09C012 homo sapien
32	38	50.7	1292	10	09L650	09L650 oryza sativ
33	38	50.7	1475	5	09N4G4	09N4G4 caenorhabdi
34	37.5	50.0	676	4	096N18	096N18 homo sapien
35	37	49.3	106	16	098377	098377 rhizobium l
36	37	49.3	147	15	09QP21	09QP21 human immun
37	37	49.3	190	3	09U073	09U073 schizosacch
38	37	49.3	251	17	097VB2	097VB2 sulfolobus
39	37	49.3	258	2	09X4J0	09X4J0 streptomyce
40	37	49.3	263	10	0943B5	0943B5 oryza sativ
41	37	49.3	359	10	P93794	P93794 tritlicum ae
42	37	49.3	360	4	09P2Y6	09P2Y6 homo sapien
43	37	49.3	373	10	022116	022116 tritlicum ae
44	37	49.3	387	10	09PE02	09PE02 tritlicum tu
45	37	49.3	395	17	097WP1	097WP1 sulfolobus

## ALIGNMENTS

RESULT 1  
ID 062730 PRELIMINARY; PRT; 2343 AA.  
AC 062730;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE FACTOR VIII.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_Taxid=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY, SPLEEN;  
RA Gordy P.W., Bowen R.A.;  
RT "Characterization of the canine factor VIII cDNA."  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.  
DR EMBL: AF049489; AAC05384.1; -.  
DR HSSP: P00451; 1CRG.  
DR InterPro: IPR001117; Cu-oxidase.  
DR InterPro: IPR00421; FA58\_C.  
DR Pfam: PF00394; Cu-oxidase; 3.  
DR Pfam: PF00754; F5\_P8\_Type\_C; 2.  
DR SMART: SM00231; FA58C\_2.  
DR PROSITE: PS01285; FA58C\_1; 2.  
DR PROSITE: PS01286; FA58C\_2; 2.  
DR PROSITE: PS00079; MULTICOPPER\_OXIDASE1; 3.  
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAAD99 CRC64;

Query Match 72.0%; Score 54; DB 6; Length 2343;  
Best Local Similarity 83.3%; Pred. NO. 0.87;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 3 SQNPPVLRKHQR 14  
Db 1648 SONPPVLRKHQR 1659

```

RESULT 2
ID 018806 PRELIMINARY; PRT; 2343 AA.
AC 018806;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FACTOR VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LAYER.
RA Cameron C., Nolley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,
RA Giles A., Lilliecrap D.;
RT "The canine factor VIII cDNA and 5' flanking sequence.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF016234; AAB87412.1; -.
DR HSSP; P00451; ICG.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_P8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;

Query Match
Best Local Similarity 72.0%; Score 54; DB 6; Length 2343;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 SONPVYLRHQR 14
DB 1648 SONPVYLRHQR 1659

RESULT 3
ID 005549 PRELIMINARY; PRT; 1077 AA.
AC 005549;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO SEVERAL DNA HELICASES.
GN YDR291W OR D9819.1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Fulton L.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Favello A., Fulton L., Galtung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier T., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
RA Teich A., Treviskis E., Vignati D., Wilcox L., Woldman P., Vaudin M.,
RA Wilson R., Waterston R.;

```

```

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Urayan, Cherry J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U51031; AAB64466.1; -.
DR SGD; S0002699; YDR291W.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1077 AA; 123548 MW; 948F024154FBE9A6 CRC64;

```

```

Query Match
Best Local Similarity 56.0%; Score 42; DB 3; Length 1077;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 5 NPVYLRHQR 14
DB 500 NPVYLRHQR 509

```

```

RESULT 4
ID 09A9F6 PRELIMINARY; PRT; 188 AA.
AC 09A9F6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC0645.
GN CC0645.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwyn M.L., Hart D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ueberback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005740; AAK22630.1; -.
DR TIGR; CC0645; -.
DR InterPro; IPR002577; DUF24.
DR Pfam; PF01638; DUF24; 1.
DR ProDom; PD004032; DUF24; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 188 AA; 21013 MW; DA14A5973FEC3B31 CRC64;

```

```

Query Match
Best Local Similarity 54.7%; Score 41; DB 16; Length 188;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 SONPVYLRHQR 13
:: :::::

```



Db 143 AEGPVLRRHK 153

```
RESULT 5
Q9PMN6 PRELIMINARY; PRT; 450 AA.
AC Q9PMN6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE N-HYDROXYCINNAMOYL/BENZOTRANSFERASE-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kottani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones."
RL DNA Res. 4:401-414(1997).
DR EMBL: AB008264; BAB09184.1; -
DR InterPro: IPR003480; Transferase.
DR Pfam: PF02458; Transferase; 1.
KW TRANSFERASE.
SQ SEQUENCE 450 AA; 50523 MW; 2325CFE8739125A2 CRC64;
```

```
Query Match 54.7%; Score 41; DB 10; Length 450;
Best Local Similarity 77.8%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 QNPPVLKRH 12
:|||||:
Db 201 KNPPVLKRY 209

RESULT 6
Q9BWE3 PRELIMINARY; PRT; 330 AA.
AC Q9BWE3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DDAI SPECIFICITY PHOSPHATASE 11 (RNA/RNP COMPLEX 1-INTERACTING).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE, RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC000346; AAH00346.1; -
DR InterPro: IPR000387; TYR_phosphatase.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
KW Hydrolase.
SQ SEQUENCE 330 AA; 38896 MW; D69TB85F6984F8F8 CRC64;
```

```
Query Match 53.3%; Score 40; DB 4; Length 330;
Best Local Similarity 63.6%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 FSQNPVLKRH 12
||:|:|:|:|
```

Db 267 FSENPHVQRRH 277

```
RESULT 7
ID 075319 PRELIMINARY; PRT; 330 AA.
AC 075319;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PIR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98352073; PubMed=9685386;
RA Yuan Y., Li D.M., Sun H.;
RT "PIR1, a novel phosphatase that exhibits high affinity to RNA.
RT Ribonucleoprotein complexes."
RL J. Biol. Chem. 273:20347-20353(1998).
DR EMBL: AF023917; AAC39925.1; -
DR InterPro: IPR000340; DS_phosphatase.
DR InterPro: IPR000387; TYR_phosphatase.
DR Pfam: PF00782; DSPC; 1.
DR SMART: SM00012; PTPC_DSPC; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
KW Hydrolase.
SQ SEQUENCE 330 AA; 38939 MW; 0C397F43043B450A CRC64;
```

```
Query Match 53.3%; Score 40; DB 4; Length 330;
Best Local Similarity 63.6%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 FSQNPVLKRH 12
||:|:|:|:|
Db 267 FSENPHVQRRH 277

RESULT 8
ID 059697 PRELIMINARY; PRT; 423 AA.
AC 059697;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE C32C12.03C (EC 2.7.1.-).
GN SPC32C12.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Badcock K., Churcher C.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: STRONG, TO S.POMBE KIN1.
DR EMBL: AL022172; CAA18163.1; -
DR HSP; P24941; ICKP.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
ATP-binding.
```

FT DOMAIN 53 305 PROTEIN KINASE.  
 FT NP\_BIND 59 67 ATP (BY SIMILARITY).  
 FT BINDING 82 82 ATP (BY SIMILARITY).  
 FT ACI\_SITE 175 175 BY SIMILARITY.  
 SO SEQUENCE 423 AA; 47673 MW; FDDDB5B532A78B80 CRC64;

Query Match 53.3%; Score 40; DB 3; Length 423;  
 Best Local Similarity 77.8%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 PPVLRHQR 14  
 ||| ||||  
 Db 393 PPVYSRHQR 401

RESULT 9  
 ID 096BV0 PRELIMINARY; PRT; 561 AA.  
 AC 096BV0;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 62.3 KDA PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA, AND CHORIOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC015152; AAL15152.1; -  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SO SEQUENCE 561 AA; 62328 MW; 505BE70818921EF CRC64;

Query Match 52.7%; Score 39.5; DB 4; Length 561;  
 Best Local Similarity 71.4%; Pred. No. 78;  
 Matches 10; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 1 SFSQNPVLRHQR 14  
 |||| | ||||  
 Db 167 SFSQKPN-LARHQR 179

RESULT 10  
 ID 09U3V4 PRELIMINARY; PRT; 225 AA.  
 AC 09U3V4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE TERRASPANIN.  
 GN TSP42EA OR BCDNA:GH05668 OR CG18817.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Page-McCaw A.W., Tsang G., Rubin G.M.;  
 FT "Sequencing Drosophila cDNAs related to tetraspanins."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF220041; AAF23825.1; -  
 DR FLYbase: FBgn0029508; Tsp42EA.  
 DR FlyPro: iPR000301; Transmem\_4.  
 DR Pfam: PF00335; Transmembrane4; 1.  
 DR PRINTS: PR00259; TMFOUR.  
 SO SEQUENCE 225 AA; 25897 MW; 80620699064248ID CRC64;

Query Match 52.0%; Score 39; DB 5; Length 225;  
 Best Local Similarity 66.7%; Pred. No. 36;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 6 PPVLRHQR 14  
 ||| ||||  
 Db 157 PPVLRHQR 165

RESULT 11  
 ID 09FTW3 PRELIMINARY; PRT; 291 AA.  
 AC 09FTW3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE EMBL|CAB88415.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=99156233; PubMed=10048488;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:379-391(1998).  
 DR EMBL: AB016877; BAB11631.1; -  
 SO SEQUENCE 291 AA; 32563 MW; 5F9B56F48C37F5CD CRC64;

Query Match 52.0%; Score 39; DB 10; Length 291;  
 Best Local Similarity 66.7%; Pred. No. 48;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 4 QNPVLRHQR 12  
 |||| | ||||  
 Db 75 ENPVLRLH 83

RESULT 12  
 ID 09W0F6 PRELIMINARY; PRT; 445 AA.  
 AC 09W0F6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE CG9169 PROTEIN.  
 GN CG9169.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goez J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Ffannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.D., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibbegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacled J.M.,  
 RA Palazzolo M., Plutman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtsek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003471; AAF47492.1;  
 DR FJBASE: FB90035217; CG9169.  
 SQ SEQUENCE 445 AA; 51459 MW; 3F6CB46D7424712E CRC64;

Query Match 52.0%; Score 39; DB 5; Length 445;  
 Best local Similarity 42.9%; Pred. No. 75;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 1 SFSQNPVLRKRR 14  
 DB 16 SFRNPDLQQRQ 29

RESULT 13  
 O97VPI PRELIMINARY; PRT; 506 AA.  
 AC O97VPI.  
 DT 01-OCT-2001 (TREMblrel. 18, Created)  
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN SSO2575.  
 GN SSO2575.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=2133296; PubMed=11427726;  
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Amey M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Etrauso G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Heikamp H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doellittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.,  
 RT "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL: AE006854; AAK42702.1;  
 DR MEROPS; M61.001;  
 DR InterPro; IPR001478; PDZ.  
 Pfam; PF00595; PDZ; 1.

DR SMART; SM00228; PDZ; 1.  
 DR PROSITE; PS0106; PDZ; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 506 AA; 58757 MW; AB084C80F24FD103 CRC64;

Query Match 52.0%; Score 39; DB 17; Length 506;  
 Best local Similarity 60.0%; Pred. No. 86;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 FSNPVPYLR 11  
 DB 379 FSRNPPIFR 388

RESULT 14  
 O94400 PRELIMINARY; PRT; 563 AA.  
 AC O94400;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE AT5G41990/MTC20\_9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
 RA Bohn J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Mirada M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,  
 RT "Arabidopsis cDNA clones,"  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF24629; AAL11622.1;  
 SQ SEQUENCE 563 AA; 63802 MW; 0342EE9FDB5A0F54 CRC64;

Query Match 52.0%; Score 39; DB 10; Length 563;  
 Best local Similarity 77.8%; Pred. No. 96;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 SQNPVYLR 11  
 DB 148 SQNPVYLR 156

RESULT 15  
 O9FHY4 PRELIMINARY; PRT; 608 AA.  
 AC O9FHY4;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MAP KINASE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=99397451; PubMed=10470850;  
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,  
 RA Miyajima N., Tabata S.,  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.

RT Sequence features of the regions of 1,011,550 bp covered by seventeen  
 RT p1 and TAC clones.":  
 RL DNA Res. 6:183-195(1999).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AB017067; BAB08432.1; -.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00069; kinase; 1.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR SMART: SM00219; TYKc; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_SF; 1.  
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 KW  
 SQ SEQUENCE 608 AA; 69479 MW; C2AB8CF51693CFD5 CRC64;

Query Match 52.0%; Score 39; DB 10; Length 608;  
 Best Local Similarity 77.8%; Pred. No. 1e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 SQNPPVLKR 11  
 |||||: |  
 Db 193 SQNPPVILHR 201

Search completed: June 13, 2002, 12:15:56  
 Job time: 219 sec